own experiences in the data gathering process as data that are part of the whole phenomenon being studied. Stated differently, the reactions of the studier are considered part of the information available about the studied. These spontaneous observations about the observational process may include perceptions of the setting and of the self as observer, and they may be incorporated on the spot to inform the ongoing activity from which further observations are generated. This kind of integration of the observer into what is observed is most clearly illustrated in participant observation and ethnographic research in which the investigator joins in the social life of a group in order to learn more about it. While an investigator in such a tradition is free from preplanned constraints on how data are to be gathered or how observations are to be made, the responsibility for the intensive self-examination that is required is great. Continuously documenting this self-reflective process is essential through diaries, field notes, memos, and other techniques that encourage reflection not only on the content of the study but on the process of it as well (Padgett 1998).

The Political Context of Research

Except in the field of evaluation research, the political context of research is an aspect of the broader context of data collection that is rarely discussed except by those with a critique to make of the status quo. There are always various stakeholders in any piece of research, however, and these stakeholders may have similar or different interests. Some of these interests may affect aspects of the data collection process from access to a site to selection of the questions to ask to selection of who will do the asking, as well as the uses to which the results of a study may be put when it is done.

Conducting any data collection effort requires resources, money, or the equivalent in time and effort. In addition to material resources, intangible resources, such as sanction, encouragement, and moral support, are often needed as well. Whoever provides the resources for a piece of research usually has a stake of some kind in its outcome. This stake may be in one aspect of the effort rather than in any one specific outcome or finding. However, it is important to know about any piece of research who the stakeholders are, especially if they are the sponsors or others who provided resources.

When views on a subject are polarized, people sometimes think that scientific or research-based information can be used to "settle" the dispute, but few findings in research are so unequivocal or open to only one interpretation that such a strategy will work. In conducting any data gathering effort, it is essential that any political conflicts among those designing and conducting the research be identified, aired, and resolved or set aside. For example, when two groups of people have an interest in the findings of an interview study but they disagree about some aspect of the problem being studied, it is not wise to have members of either group conduct the interviews because members of one group may not trust the information provided by members of the other. Getting them to agree on someone whom they both trust to gather the data and on how the data will be gathered will ensure that the findings from the interviews will be accepted and used.
Finally, it is important to anticipate the political uses, positive or negative, intended or unintended, to which any findings can be put before data collection is begun. Often there is little control that can be exercised after the fact over how and by whom research results will be used. As far as is possible, then, representatives of any groups, organizations, or communities that may be affected by the anticipated findings should be consulted before data collection begins (Bowman 1991).

Selecting a Research Setting

All of the considerations discussed so far in relation to the context of research have been fairly abstract. The selection of a setting in which to conduct a study and in which to gather data, however, is also a very practical one. The following set of questions reflects both the theoretical and the practical dimensions that must be considered when choosing a setting for research.

What context, setting, or situation(s) will be most likely to allow for the phenomena of interest to be seen or to emerge? This question comes first because it is by far the most important one. Whatever the method of data collection to be used when observing or interacting with people, the setting in which they will be seen or spoken with may have a great impact on what they do or say and thus on what can be observed in the research. For example, if the focus of an observational study is on how residents of a nursing home do or do not interact with each other, it might make more sense to observe people in the common or dining rooms rather than in their individual bedrooms or bathrooms because those are the places in which they are most likely to encounter each other. However, to the extent that residents are accustomed to visiting each other and socializing in their own rooms, that behavior would be lost to study. In an interview study of marital satisfaction, for example, interviewing couples at home or in an office, alone or in the presence of others, together or separately, are all likely to affect the kind of information that is shared.

Will the setting contemplated allow for the kind of data collection needed to be done? Will it allow for the kind of recording of information that is needed as well? Once the kind of setting that will best allow the phenomena of interest to emerge has been identified, it is also important to consider how well the setting will accommodate the kind of data gathering that is contemplated. For example, when studying the social interactions of nursing home residents through observation, the common rooms might most easily accommodate that form of data collection; in an individual room, much time might pass without any interactions to observe. If, on the other hand, the plan is to interview the residents about their social lives and their satisfaction (or dissatisfaction) with them, it might be better to conduct the data collection in rooms with greater privacy, perhaps leading to greater candor. It might also involve less background noise, making hearing the questions and answers easier for participants and interviewers.

The setting chosen for data collection must also be able to accommodate whatever means of data recording that will be used if the recording is taking place on
site. Traditional methods of participant observation research, for example, have involved the generation of extensive field notes after each period of time spent in the setting. If such a method is to be used, the periods of data collection on site must be planned in such a way as to allow for periods of recording directly afterward.

When the recording of observations will take place on site simultaneous with the data collection itself, planning must allow for that as well. If videotaping or audiotaping will be used, will the data collection setting accommodate it? And how will it be explained to and experienced by participants and others in the setting? Paper and pencil recording may often seem less intrusive, but even the presence of a person with clipboard, paper, and pencil in hand may have a meaning and impact in the setting that should be anticipated and addressed. Finally, in some cases portable computers are now used for recording information on site; again, if used, a data gathering setting that can accommodate this method must be selected.

Is the setting contemplated for the data collection accessible? Can the researcher count on being able to gain access to the site? This question is one both of can and of may. The “can” part speaks of feasibility for the researcher; the “may” part speaks to consideration of ethical issues and of the permissions that may be necessary for legitimating the conduct of the data gathering effort. To return to the example of the nursing home study, a researcher is likely to select a nursing home that is geographically convenient in which to conduct the research. This strategy makes sense; energy can then go into observation itself rather than into travel to the site.

Once the nursing home has been selected, however, although an observer can most likely enter a nursing home’s common areas to make observations under the guise of being a visitor to a resident, this strategy would not be ethical or prudent. The permission of those who own or have responsibility for the setting in which any data collection will take place should be sought not only for ethical reasons (see chapter 9) but also to legitimate the research. There are very few spaces that are truly public. In conducting research, as in any other professional encounter, sanction for the activity based on its purpose is very important to obtain. The choice of specific sites for data collection, then, are often based on whether or not legitimated access to the facility or setting for research purposes can easily be obtained or not.

How can the setting be described so that it can be compared with others that are similar and distinguished from others that are different? How representative is the setting of others that might be assumed to be like it? An area that is often neglected in reports of research studies is how the sites for the data gathering were chosen and what they were actually like. The issue of how well a given sample of individual study participants represents some larger population is commonly considered, but the fact that a given setting may (or may not) represent some larger population of possible settings is not often discussed. Except in large scale survey studies, settings are rarely sampled in the way that individual participants are. Even so, a description of the setting and of the data gathering context is necessary for the reader of a study to determine how applicable the results might be in other contexts or how the context of the data gathering may have affected
Basic Issues in Data Collection

what was learned. All phenomena are studied in a context, and the context influences what is observed. Therefore, describing the contexts of any data gathering effort is essential.

How to Collect Data

There are three basic methods used for the collection of original data in research: observation, interviewing, and written self-report. Each of these has a long tradition of use in the social sciences and in human service research, and each has its advantages and drawbacks that make it better suited for gathering some kinds of data than others. While these advantages and drawbacks are described in depth in the chapters devoted to each method, an overview of them is given here.

Observation

The term observation has two uses in the literature on research: the process of experiencing and recording information for research in general based on some kind of sensory experience of it and, more specifically, the process of gathering visual information. As a specific data collection method in research, observation is currently underutilized in social work research despite the fact that when appropriately used the data that flow from it are often very convincing.

Observation as a specific data collection method covers a range of activities. At one end is the passive, noninteractive process of watching—and perhaps also listening—to what is transpiring in the research setting. At the other, the researcher might take a much more active stance by participating in some interaction or performing some planned intervention and then watching—and perhaps also listening to and experiencing in other ways—what transpires before and after the intervention. What is observed may be carefully and precisely predefined prior to the data collection process, or it may be only loosely defined to begin with and then further articulated by the observer once in the setting itself or as a product of the data analysis.

Observation as a data collection method is often best suited to the study of behavior, observable actions, and interactions. What people say they do and what they actually do can differ. Direct observation is often the most effective way of learning about people’s behaviors, individual or interactive, as distinct from intentions, motivations, or the recollections of them. Jarrett (1992) (see chapter 4) used observation effectively in her case study to ascertain who really was involved in the family life of one household, discovering that many more people participated regularly in it than data on residence would have suggested. However, some behaviors are normally very private and thus difficult to observe. There are often limitations to learning about intimate or illegal behaviors, for example, through direct observation.

Observation has traditionally been termed a reactive method of data collection. Reactivity is the tendency to evoke a response in those being studied to the data gathering process itself. This reactivity arises from the fact that unless the
observational method involves concealing the process from participants, as through one-way mirrors, hidden cameras, or covert observation, the observer(s) and perhaps also observational equipment, such as cameras, are on site, visible, and known to those being observed. Those being observed, therefore, are assumed to have reactions to this presence, and those who observe must acknowledge this potential and react to it in turn. How an observer is introduced into a setting, how the observer appears, what explanation for his or her presence is given, what kinds of interactions do and do not take place during the process of observation, how the observer reacts to questions or challenges about the role, and how the observer leaves the setting—all of these factors are likely to affect the kinds of reactions that those in the setting will have to the data gathering process itself and should be considered when the data collection is planned and when the data are analyzed.

Observation as a data collection method lends itself to the collection of either structured—specific and quantified—or unstructured data. There are protocols in existence to guide the collection of observational information about specific phenomena in a predetermined way. There is also an extensive literature on participant observation and ethnographic research and the techniques used to gather observational information in an unstructured way. In this tradition, observation is often combined with interviewing, with the questioning of participants about events observed, reflecting the flexibility of data gathering in these traditions. Observation, then, while especially useful for the study of nonverbal behavior, can also embrace the study of verbal behavior as well.

**Interviewing**

Interviewing is probably the most common method of data collection used in social work research. An interview is a conversation conducted for the purpose of eliciting information from someone else. As with observation, interviewing can be used to collect data that are either structured—specific, predefined, and quantified—or unstructured, also called narrative. Interviews can effectively be done both face-to-face or over the telephone. However, because an interview is a verbal medium, a sophisticated understanding of all the factors that influence verbal interactions is essential (Foddy 1993).

Interviewing as a data collection method is focussed on verbal behavior, on the words used by people to describe events, recollections, opinions, attitudes, feelings, motivations, intentions, and meanings. These are essential aspects of people’s psychological and social lives that in fact cannot usually be directly determined, but only inferred, from behavior; they are aspects of another person’s internal and interpersonal world that often can best be apprehended through their words, through their own verbal description of them. A person’s face may express sorrow; tears may be seen to flow. In most cases, however, the explanation of why the person is crying or what the feeling being expressed through the tears is can only be described based on the person’s answer to a question about it. Written accounts may also be used to capture the words that people use to describe these internal experiences—feelings, meanings, attitudes, and the like. However most people find it easier to express what they mean orally rather than in written form,
Basic Issues in Data Collection • 309

especially when the answers or descriptions sought are open-ended or structured
only by the participant or, in some cases, when the information sought is of a
complex or sensitive nature. However, transcripts of research interviews or other
verbal interactions, such as treatment encounters, can be valuable data sources for
research.

Interviewing is used in social work and other helping professions all the time
for purposes other than research. The knowledge that most helping professionals
have about interviewing can provide a foundation for research interviewing,
which is similar in some ways and different in others from interviewing for other
purposes (Bunin et al., 1983). Skills in establishing rapport, in asking questions
that can further illuminate what has already been said, in conveying an accepting
and nonjudgmental attitude toward what is said that encourages further explana-
tion, and in modulating personal reactions to what is said—all of these are assets
in most research interviewing situations. However, since the purposes of the data
gathering are different in research than in the helping situation, adaptations must
be made to keep the interview process focussed on the aims of the research: to
gather in a structured or unstructured fashion information limited to answers to
the general research question that guides it.

Interviewing, like observation, is a fairly reactive method of data gathering.
When interviewing, especially in face-to-face interviews, the observer and the
participants in the research encounter each other directly and react not only to
what is said but to how it is said, to who the interviewer is, to how he or she looks,
and to how he or she behaves. Even in telephone interviews, assumptions may be
made about who the interviewer (or participant) is and what he or she intends
based on the sound of the voice rather than just on what is explicitly said. Specific
behaviors, aspects of appearance, and phrasing and tone of voice can be
consciously used in the interview situation to aid in the effectiveness of the data
gathering by establishing a positive rapport between interviewer and participant.
On the other hand, some reactions and responses cannot easily be anticipated.
When planning an interview study, therefore, careful consideration should be
given not only to the explicit content of the interview encounter, to the questions
that will be asked and the explanation that will be given for them. It is also impor-
tant to consider the other aspects of the encounter that may naturally affect the
kind of data that are gathered.

Written Self-Report

Written forms of data collection are also very common in social work research,
especially in the form of mailed surveys or questionnaires. In addition, standard-
ized measures of specific psychological, attitudinal, and other characteristics exist
that are often incorporated into social work research as the Gomez (1990) study
illustrates. This method of data collection is distinguished from others by the fact
that the research participants contribute by writing down the needed information.
Although this occurs most often by checking off fixed, predetermined responses
to preformed questions, written data that is narrative in form can be collected as
well, as in journals, logs, or answers to open-ended questions.
The ability of anyone to participate in research that relies on the collection of written data depends, of course, on being able both to read and write in the language of the materials prepared. This factor can be a major one limiting the use of this form of data collection. In addition to those with limited literacy or without knowledge of the language being used, children and, in some cases, those with visual or motor problems also may not be able to participate effectively in providing written data. Sometimes when open-ended data are needed for interpretation, children may be asked to express themselves through drawings—a specialized and interesting form of data collection in itself.

In addition, a researcher wishing to collect verbal data in written form is generally limited to using the languages that he or she knows well. Although the problem of having a common language can affect interviews and even observation as well, there are people who can communicate well orally in a given language who are not as comfortable or effective in written communication in it and thus who might have difficulty with a questionnaire.

One of the great advantages of collecting data in written form is that, if needed, it can be done anonymously, that is, without any direct contact between those who gather the data and those who provide it. The mailed questionnaire or the survey voluntarily picked up, filled out, and dropped in a box are examples of anonymous data collection procedures. Mailed surveys, of course, can be identifiable, and it takes a special effort to keep them anonymous. However, the written self-report is the only form of data collection in which complete anonymity is even possible because the participant is not seen or heard and cannot be located through the response. Although the telephone interview may seem anonymous, the identity of the interviewee can often be determined through knowing the telephone number that has been dialed; in the truly anonymous situation the identity of the participant is both unknown and unascertainable.

The potential anonymity of written data collection means that it is often used for topics that are sensitive or that address illegal, socially sanctioned, or intimate behavior, when it is believed that participants will disclose more in an anonymous situation than when they are identifiable. Information about drug use or sexual behavior, for example, is generally collected in this way.

In addition, although it is possible to use written forms of data collection for narrative data or unstructured responses, because writing at length is generally difficult and time-consuming, data collected in written form are usually prestructured and fixed in form. This requires, of course, that considerable effort be devoted in advance to developing or selecting the questions and response categories that constitute the survey or other data collection form. Unlike in observation or interviewing, then, there is generally little flexibility to the data gathering process once it is underway. Therefore the collection of data in written form is generally associated with fixed rather than flexible method research. However, existing written materials are often incorporated into flexible method studies (Padgett 1998).

One of the major questions to be answered when collecting written data in fixed method research is whether to develop an instrument for this purpose that is unique to a given study or whether to identify and use existing instruments to
measure the concepts of interest in the study. With a data collection tool developed specifically for the study at hand, the nature of the data to be gathered can be tailored exactly to the purposes of the study and to the ways of conceptualizing and defining the phenomena of interest that seem most appropriate. However, because of being new and unique, these methods of eliciting data will not have been previously examined for their effectiveness. Much effort will be required to refine and pilot the questions and response categories developed, and flaws that lead to incomplete or ambiguous information are often encountered. For these reasons, researchers often seek out "tried and true" instruments in order to benefit from the prior work of others in developing and debugging easily interpretable measures (see chapter 15). Use of an existing measure, however, requires that the phenomenon it purports to capture is in fact the same as the concept or phenomenon of interest as defined in the present study, and it may or may not be easy to identify an existing instrument that fits the study well. However, there are many tools available to assist in identifying measures that exist and evaluating them for use in a particular study or with a particular type of participant.

Other Aspects of How

So far, the discussion of how to collect data has focused on the mode of data collection—observation, interview, or written self-report—to be used. There are other questions about "how" to answer, however, especially when structuring questions in fixed method research. In most cases, the goal in gathering data for fixed method research is to be as direct, precise, and unambiguous as possible about the information that is needed when asking a question. Yet there are times, such as when exploring complex feelings or motivations, that some ambiguity in a question is desirable in order to give the participant the scope to answer it in a variety of ways or in the way that best describes a complex response or situation. In addition to structuring the question, of course, there are the problems of how and how much to structure the response. In general, prestructuring the responses or types of observations to be made increases the comparability of the data recorded for or by each person. It forces the participant or the observer to reduce the natural variation in expression to the categories that exist. The chapter on questionnaire design (chapter 14) later in the book goes into some detail on the choices to be considered in prestructuring both questions and responses, concepts that pertain to interviewing and observation as well in the context of fixed method research.

Who Will Collect the Data

Along with consideration of the content to be covered in the data gathering effort, the context in which the data might best be gathered, and the method of data gathering that should be used, the question of who might best provide the information that is needed should also be considered. Asking this question does not intend to open the question of the best sample to employ in the research (see chapter 10).
Rather it asks, given the nature of the sample, who might best provide the information needed about the people in it: the people themselves or someone else? It also asks, especially in fixed method research, if data are to be gathered from sample members in person, what characteristics and training should this person or people possess?

The question of who might best provide the data is a question of choosing an informant. Suppose, as in an earlier example, there were a study being planned to examine the social activities of a group of nursing home residents. There are several possibilities for informants in such a study. These informants might include the residents themselves about themselves, the residents themselves about each other, staff in the nursing home, visiting friends and family members of the residents, or research observers with no other relationship to the residents or the staff.

Each of these potential informants provides information that is conceptually distinct. All can potentially provide information about behavior either as directly observed, as summarized and reported verbally, or (the residents themselves) through self-report. All can provide information about the motivations for that behavior or the feelings accompanying it, most as inferred but only some, the residents themselves, as directly experienced. Some groups, such as the residents themselves, may have functional characteristics that potentially limit their usefulness as informants, such as perceptual or memory problems. Some may be able to provide written information if that is needed; some may not. All of these potential informants will have a particular point of view on the study issue: Staff and family members, for example, may differ in their perceptions of the adequacy of the opportunities for socializing that the nursing home provides, which may in turn be reflected in the information about residents’ social behavior that is furnished to the study.

The choice of an informant for such a study must depend on all these factors: what the specific data needed are, how they can best be collected, what form they will be collected in, the capacity of the various potential informants to provide it, and the points of view that the various informants might bring to the process. The most important determinant of the choice must be the nature of the phenomena to be studied, what the specific data are that are needed. The other choices—of informant and of ways of collecting the data—are made in order to get the specific information needed in the most direct and efficient manner possible. Often trade-offs are made; for example, even though data from residents directly might be preferable, if the group to be studied includes many frail or sick individuals, using other informants might be the most feasible way to proceed although the point of view of the residents themselves would be lost.

When resources are adequate or the decision about informants is difficult to make, multiple informants are sometimes used. This strategy often affords excellent opportunities to understand a phenomenon from more than one perspective and to assess it in more than one way, resulting in a fuller picture of a phenomenon. In studying children, for example, it is commonly acknowledged that parents and teachers often do not describe the same child in exactly the same way but that both often have valuable information to contribute about the child’s
behavior and functioning. When using multiple informants and when their perceptions differ, it can be difficult to determine how to reconcile their accounts of the phenomena of interest. Conversely, when the reports of different informants agree, the data obtained tend to gain in credibility.

When data are collected and the findings of a study are reported, it is essential that the data be described in a way that makes the nature of the informant clear. For example, data from this hypothetical nursing home study might be described as “nursing home residents’ self-reported patterns of socialization” or as “staff perceptions of residents’ socialization patterns,” depending on who provided the data. While the content of the study in these two cases is in some respects the same—residents’ patterns of socialization—the basis for describing it is quite different and must be made explicit. This procedure is one way in which an aspect of the observational context is commonly made clear in reporting the results of fixed method research.

Subjectivity and “Objectivity” in the Data Gathering Process

Choosing the informant for a study is one way of determining whether subjective or “objective” data will be used. By objective data in this instance is meant only that the data are furnished by someone different from the one having the experience of primary interest. Returning to the nursing home residents, a resident’s own account of his or her own socialization is subjective; someone else’s account of the resident’s socialization is not and may be termed objective by contrast. In this sense, neither of these two forms of data is any more inherently trustworthy (or not) than the other. A person may try to deceive others about his or her own activities or may have engaged in self-deception about what he or she really does; similarly, an observer may consciously or unconsciously perceive, or report on his or her observations of, another accurately or not. In this sense, subjective data and objective data are neither inherently better nor worse; they are simply different from each other.

There is another sense, of course, in which the terms objective and subjective are used evaluatively when describing data. Subjective data are often defined as existing only in the mind, without reference to external reality and not trustworthy for that reason. There are some phenomena of interest in social work research, such as feelings and other internal psychological states, such as motivations, that are inherently considered to exist by definition only in the mind. The challenge is to find ways to convey and record such subjective phenomena, by definition not directly observable, for the research. By contrast, however, the term objective can mean both having actual existence in reality and “being uninfluenced by emotions or personal prejudice” (American Heritage Dictionary 1992:1247). It can thus mean observable, as objects or behaviors are in contrast to emotions, for example. However, it can also mean unbiased, that is, a description as close to the reality experienced as possible. In this latter sense, even subjective data can be more or less objectively rendered. Much of the effort that goes into any data collection process, of course, is designed to produce information, whether about subjective or objective phenomena, that are as objective, that is, unbiased, as possible.
However, some question the idea that any data are unbiased in the sense of being uninfluenced by the standpoint and culture of the person doing the observing or providing the data (Landrine, Kolonoff, & Brown-Collins 1992).

For any observer in the research context, there is one final level on which to consider the objectivity and subjectivity of what is observed. In fixed method research, many studies and most forms of interaction between those who gather data and those who are the participants in the research are designed so that only specific verbal or behavioral data are to be noticed and recorded. In some flexible method research, however, the observer enters the research setting and enters into interaction with the participants in the research in a fuller way, through, for example, participant observation or unstructured or intensive interviewing processes. In such situations, there may be many subjective responses evoked in the observer by the data elicited or the situations encountered. These subjective responses of the observer may either be dismissed, considered as contaminants of the observational process, or considered to be part of the data, although subjective, that the observer has that might inform the interpretations made of the data. In other words, these subjective reactions in the data gathering situation may either be defined as part of the data available to the study or not. Although not commonly acknowledged, these subjective reactions to the data gathering process can occur in the context of a fixed method study as well, but they are especially likely to occur when contact between the researcher and the researched is intensive, prolonged, and/or interactive.

If such subjective reactions are to be used as data, they must be treated as such during the data gathering process. If, for example, a structured interview is being conducted, written notes can be made on the interview guide, perhaps at the end of it, recording the observer’s responses to the situation for later analysis. In intensive interviewing and participant observation research, it has long been acknowledged that considerable time is needed after each data gathering episode to record in depth the full range of responses of the interviewer or observer along with the data that were gathered about the participants.

The Use of Multiple Observers

Returning to the issue of who will collect the data, there are particular issues to be considered in data gathering when the informants for a study are in the setting only as researchers and when more than one of them is involved. The situation of multiple observers or reporters is, in fact, quite a common one in fixed method research, especially in survey research and other large-scale fixed method studies in which multiple observers or interviewers are used. The problem that this situation poses is, quite simply, that of distinguishing variations in the data collected that arise from differences in what was observed from those due to who the observer or reporter was.

As has already been noted, in most fixed method research the effort during data collection is to hold the observational context constant as much as possible throughout the process. Even when a standardized way of gathering data has been developed, the fact that different people are involved in actually doing the data
Basic Issues in Data Collection

collection is an obvious source of possible unintended variation in the data. The first strategy used to deal with this problem is to record who the person gathering the data was in each instance in a systematic way. When the data have been gathered, they can then be analyzed to see whether or not there are any systematic differences evident in the findings generated by different observers or recorders.

Before any data are collected, however, it is essential to orient and train all of the people who will be gathering data to the data collection process being used in the study, emphasizing the importance of conducting the data gathering in as consistent a way as possible. While rules to guide the data gathering are important, unanticipated situations always arise whenever data are actually being collected. One benefit of having a number of people engaged in data gathering in the field is the information they can bring back to the study about how the data collection process is actually working out in practice. In particular, they can inform those who are directing a study about the situations they encounter that do not fit well with what participants actually seem to be saying or doing or with the rules for data gathering that have been developed. It is essential, then, not only to work with those who will gather the data at the beginning of a study but to continue to work closely with them during the data collection process so that adjustments can be made across the board based on information coming from the field. Only in that way will the most consistent and trustworthy data be collected.

Evaluating the Usefulness of Data

Whatever the method used to collect data in research and whoever may provide it, in the end the ultimate question that must be asked about any data gathered is how useful they are for answering the research question(s) asked. This broad question about data collection has generally been broken down into two parts based on traditional notions in measurement theory: Are the data reliable? And are they valid? Although traditional measurement theory has depended on logical positivist assumptions that are not the same as those of fallibilistic realism, the concepts of reliability and validity, if more broadly defined, remain useful and important ones to consider when planning or evaluating a data collection effort. Together they really address the general question of how useful and trustworthy the data generated by any given study are.

Validity and especially reliability have more often received formal discussion in fixed method research than in flexible method traditions. From the standpoint of fallibilistic realism, however, the importance of both concepts rests on certain assumptions that apply to both kinds of research:

There is a world of empirical reality out there. The way we perceive and understand that world is largely up to us, but the world does not tolerate all understandings of it equally (so that the individual who believes that he or she can halt a speeding train with his or her bare hands may be punished by the world for acting on that understanding). There is a long-standing intellectual community for which it seems worthwhile to try to figure out collectively how
best to talk about the empirical world, by means of incremental, partial improvements in understanding. Often these improvements come about by identifying ambiguity in prior, apparently clear, views, or by showing that there are cases in which some alternative view works better. . . . “Truth” (or what provisionally passes for truth at a particular time) is thus bounded both by the tolerance of empirical reality and by the consensus of the scholarly community (Kirk & Miller 1986:11–12).

Concepts of reliability and validity have been essential in guiding discussion of data collection methods both in terms of how well any data tolerates a meaning or interpretation assigned to it and the consensus that is likely to be achieved in the scholarly or professional community about the trustworthiness of those data. Following Kirk and Miller (1986), reliability is defined as the degree to which a reported observation is independent of the accidental circumstances of the data collection process, and validity is defined as the degree to which the data have been interpreted in the right way. Viewed in this way, the concepts of reliability and validity clearly have relevance to data collection for research that is flexible or fixed in method. However, in constructivist traditions of qualitative research the concept of trustworthiness is often used to embrace and expand on traditional notions of both reliability and validity (Lincoln & Guba 1985; Padgett 1998).

Reliability

The concept of reliability has had much more attention in fixed method research than in flexible method traditions. Reliability as traditionally defined in fixed method research has to do with how repeatable or replicable a given observation is. In traditional measurement theory, any given observation made in research is considered to be made up of two components: “truth,” the way the thing observed “is” in some absolute way in the world, and “error,” the degree of distortion inevitably introduced by any measurement process. This inevitable error inherent in the act of measurement is assumed to be randomly distributed, to distort perceptions of it in no particular direction. Errors of perception or measurement that have a direction are defined as bias because they introduce a systematic rather than random inaccuracy into the data collected.

Reliable measurement is defined as having a high proportion of truth and a low proportion of error, either random or systematic. Reliability in measurement is generally described as repeatability or agreement. If there is a high degree of error in a measurement, no two (or more) people making the same measurement of the same thing using the same tool or data collection method are likely to come up with the same observation, or score, in quantitative terms. If there is only a small amount of error in the measurement, however, their observations are likely to agree quite closely.

In everyday life, when dealing with the physical environment, there are many illustrations of this concept of reliability that come easily to mind. Everybody knows that most bathroom scales, for example, operate using a spring mechanism
and that such mechanisms are often not very reliable. However, properly calibrated scales that use weights and a balancing mechanism are more reliable. They are much more likely on repeated use to show the same weight for the same person on the same occasion than an ordinary set of bathroom scales will. Similarly, thermometers that use a column of mercury generally give more reliable (and more accurate) measurements of air temperature in general than do thermometers that use a metal coil.

The concept of accuracy is a bit different from that of reliability. Suppose that the truth of an observation, perhaps the air temperature, can be known absolutely. The most accurate measurement is the one that comes closest to what the true temperature actually is. It is possible in theory to have a reliable measurement that is not accurate; for example, there might be a thermometer that registered an air temperature that was always exactly three degrees too high. Such a thermometer has a bias toward a high temperature. It would be considered reliable, giving a consistent reading under the same conditions, but it would not be considered accurate, that is, reflective of the true temperature. Nevertheless, the concept of reliability is generally used to include accuracy as well: A measurement is most apt to be consistent or repeatable when it is also accurate, or reflective of the true phenomenon.

Sensitivity is another characteristic of a measurement tool. A sensitive measurement is one that can detect even small changes in the phenomenon when they occur. By contrast, stability in a measurement tool is the quality of being able to show consistency when the phenomenon itself does not change across occasions of assessment. While both of these qualities are desirable in an ideal data collection procedure, they are in some ways opposite from each other. A data collection tool designed to be sensitive may not show stability well; one designed to be stable may not be sufficiently sensitive to change. The formal assessment of the reliability of any data collection procedure must be undertaken knowing that the accuracy, sensitivity, and stability of the measurement, and of the phenomenon it measures, must also be considered.

Redefining Reliability

This conceptualization of data collection as measurement clearly implies that the thing observed is separate from the observer and that there can be one “best” or most “truthful” version of reality developed. Fallibilistic realism holds that there is a real world separate from the observer, but it emphasizes the fact that all descriptions of it are a product of the thing itself, the concepts used to define and describe it, and the standpoint of the describer. This framework is not a relativistic one in which all descriptions are necessarily of equal merit or, as Kirk and Miller (1986) put it, “tolerated” equally well by real world events. Understanding how to evaluate descriptions or data collection in this framework involves understanding science as a collective and social enterprise and the role that observations play in it.

The concept of reliability is about multiple observations and/or multiple observers. Science itself, like any profession, is a social institution made up of those credentialed and viewed in any given society at any given time as the experts in specific and socially defined areas of knowledge. It is also, of course, a
set of methods generally accepted in the scientific community as the most useful for developing new knowledge. In this context, reliability in data collection can be understood as consistency or repeatability in the observation of a phenomenon based on the amount of agreement or consensus achieved or achievable among expert observers using common definitions and methods of observation. This idea is also called confirmability (Drisko 1997). This longer definition simply makes explicit some of the assumptions that in fact are embedded in reliability as defined in the traditional framework.

Trustworthiness is the construct used more often by qualitative researchers than reliability to describe the quality of data. Padgett (1998) describes three main threats to the trustworthiness of data in qualitative research: researcher bias, respondent bias, and reactivity. These refer to conscious or unconscious distortions in the information gathered that arise from the researcher, the research participant(s), and/or the interaction between them. Padgett (1998) also describes a variety of techniques that can be used in the data collection and analysis processes to reduce specific threats to trustworthiness, such as prolonged engagement, triangulation, peer consultation, member checking, negative case analysis, and creating an audit trail. However, the concept of trustworthiness and these methods of enhancing it also embrace ideas related to traditional notions of validity, discussed below.

**Traditional Types of Reliability**

In traditional measurement theory as applied in fixed method research, several different types of reliability have been defined. The most common statistical measures of reliability define it as internal consistency and express it as an alpha coefficient. This concept applies to data collection instruments that have many individual questions or items that are meant together to measure a single concept, as in scales of self-esteem, depression, and the like. The idea is that if all the constituent items that make up the scale are actually addressing the same concept or phenomenon, they will be highly correlated with one another. However, if in fact some are capturing aspects of some other phenomenon, they will not be as highly correlated with each other, and the group of items as a whole will not be as highly intercorrelated on average. Statistics such as Chronbach’s alpha, then, give an average intercorrelation among all the items that make up a single scale or unidimensional data collection tool as a measure of reliability.

This form of reliability, of course, is by definition applicable only to quantified observations of one particular kind. There are other forms of reliability of this kind such as what is termed split half reliability. This method of assessing the internal consistency of a scale consists of dividing its constituent items in half, either the first half compared to the second half in order or by taking alternating items and then comparing the scores obtained on the halves to each other. The idea is that if the items on the scale all assess the same thing, the scores derived from each half of it should be very highly intercorrelated; if they are not, it suggests that the items in the halves may each be assessing different things. Again, this form of reliability has application only to specific forms of quantitative data collection: the development and use of multiple item scales.
The concept of test-retest reliability, developed originally in the field of psychometrics, has more general application. This form of reliability can be generally defined as consistency in observations made of the same person and/or phenomenon on two occasions, at two different points in time. The assumption is, of course, that the person or phenomenon itself will be unchanged with respect to the characteristic(s) in question despite the passage of time and that the conditions and context of the observation are also repeatable. If these things are true, then consistency of observation over time is considered strong evidence of the reliability of a reported observation.

While the concept of test-retest reliability was developed and articulated in terms of fixed method research and quantitative measurement, it has its place in some flexible method research as well, although not formally described as such. In some forms of flexible method research like participant observation, observation may be conducted over some period of time. In selecting the most salient features of the many phenomena encountered to describe, one criterion that is often used, explicitly or not, is that the phenomenon be observed consistently, that is, on more than one occasion, if it is to be taken as characteristic of the situation or the group. Although not quantified or quantifiable, such a phenomenon can be said to have been consistently and recurrently observed, also called confirmability (Drisko 1997). That is why in qualitative research prolonged engagement is often cited as a way to enhance the trustworthiness of data (Padgett 1998).

In the traditional framework, there is a major principle affecting test-retest reliability that must be remembered: The longer the interval between the occasions of data collection, the lower the consistency, or reliability, of results is likely to be. This principle expresses the fact that even when a salient and stable phenomenon has been identified and observed, the real world is a constantly evolving context in which things observed are always changing. This is especially true of phenomena or observations at the extremes; it has repeatedly been observed that extreme states and observations are even less likely to repeat themselves than are others, a concept known statistically as regression to the mean.

A final and major form of reliability as traditionally defined is termed inter-rater or interobserver reliability. This form of reliability makes explicit the idea of consensus in observations among different people making observations or judgments about what has been observed. This consensus assumes, of course, that the content and conditions of the observations made are the same, that all that differs is the individuals making the observations. If a thing exists in the real world and if its presence or the amount of it present can be assessed by an observer using a specific definition and making certain observations in a specific way, then if another person using the same definition and procedures observes the same thing, the original observation can be said to be reliable. Because there is always some variation, traditionally termed error, in any observation process, the two or more observations are not necessarily expected to be exactly the same; rather they are expected to be similar or highly correlated with each other. If they are not, there is no way to determine which observation, if any, is the more accurate one; if they are similar, however, each separate observation enhances the reliability of the other.
Part III: The Elements of Research Design

The idea of interobserver reliability is more general than some other types of reliability, like those that apply only to quantitative scales. In participant observation research, for example, concepts or observations obtained from more than one informant in the setting or noted by more than one person involved in the research in the setting may be said to have been repeated. However, flexible method research is less likely to involve multiple investigators or data collectors than some forms of fixed method research. In addition, to the extent that observations may be the result of complex social interactions among specific actors at a particular historical moment or in a particular context, they may not be repeatable at other times or with other participants and observers. While this lack of repeatability may not negate the existence of the first observation, it does speak to the generality of the phenomenon or observation that was made. Interobserver agreement is sometimes used when analyzing records from flexible method research; when more than one observer of the record, or coder, can agree that a concept or theme appears in the data, the credibility of the conclusion may be enhanced. Hence the concepts of triangulation in data collection and peer consultation in both data collection and data analysis are cited as ways to enhance the trustworthiness of data in qualitative research (Padgett 1998).

Finally, the comprehensiveness of the documentation of the data collection process may help to support the believability of the observations made and reported (Kirk & Miller 1986). Even if never repeated, documented observations may be seen as more credible because they show how the data were generated in the first place or how variations in them can be explained. In addition, the observer/reporter will gain in credibility with other scholars simply by making the basis of the conclusions drawn as clear as possible. Padgett (1998) and others refer to this technique as creating an audit trail.

Validity

Validity has been another essential concept in measurement theory. It addresses another necessary part of scientific observation: the concepts by means of which phenomena are identified or named and then connected to theory. Validity means that the tie between evidence, sensory data, and concept or theory has been defensibly made. Reliability addresses what was seen or observed; validity addresses the meaning attributed to the observation. Simply put, validity is calling things by their right names (Kirk & Miller 1986).

As already noted, the traditional view of research has seen concept and evidence as wholly separable and as tied together by an operational definition derived from a conceptual one. Fallibilistic realism, on the other hand, also views evidence itself as shaped by concepts. Both the traditional point of view and fallibilistic realism recognize that the “fit” between evidence and the concepts used to explicate that evidence can be better or worse. Evaluating validity therefore depends on examining the correspondence between specific observations and either different observations thought to represent the same concept or, simply, the concepts themselves. As with reliability, however, fallibilistic realism reminds us
that both the perception of what counts as data and the definition of the concepts of interest are the product of consensual processes generally confined to some community of experts that is socially defined. In assessing validity, then, both the data and the meanings assigned to them are matters of discussion and agreement whether explicit or not.

In fixed method research, the assessment of validity involves assessing the operational and conceptual definitions of the phenomena studied to see how well they correspond. It may also involve identifying multiple types of evidence that can be considered to identify the same concept or phenomenon and then examine how well these multiple indicators of the same thing cohere. In flexible method research, while formal, preformed operational definitions do not exist, the examination of validity tends to proceed in two different ways. One way is through the grounding of conclusions drawn and the interpretations made in a convincing degree of detail from the data, often detailed enough that the reader can draw independent conclusions about them. It can also involve consulting with research participants about how the data should be interpreted, a technique called member checking (Padgett 1998). Sometimes, as in fixed method research, the trustworthiness of data can be enhanced by seeking multiple instances or indicators of the same phenomenon, called triangulation (Padgett 1998). It is also done by contrasting it with different phenomena, as in negative case analysis and other ways of seeking out potentially disconfirming data (Drisko 1997; Padgett 1998).

Traditional Types of Validity

In fixed method research, the most common way of examining validity is by assessing face validity. Face validity simply considers whether the manifest content of a data collection instrument or question actually seems to address the concept used to label it. It asks the question, “On the face of it, does this question really seem to capture the concept intended?” A question about caregiving that asks “How many times per week do you prepare meals for your mother?” would be considered to have greater face validity than “Do you love your mother?” or even “How often do you visit your mother?” The latter questions might or might not be relevant to caregiving, but the first question clearly is. It is said to have face validity because the content of the question clearly fits well with the general concept that the question is designed to capture.

Notice that a question does not intrinsically have face validity or not; it does or does not have face validity in relation to the specific concept that is being examined at the time or to the specific context in which it is being asked. The question above about visiting, for example, might have great face validity in a study of social contacts and support among adult family members even though it may not be the best one to ask when caregiving specifically is the focus. Questions without face validity can be considered intrusive, offensive, or deceptive by those who are asked to answer them. Therefore, it is important to consider face validity when asking questions in any form of inquiry, whether flexible or fixed in method.

The apparent face validity of a question can be deceptive. For example, Landrine, Klonoff, and Brown-Collins (1992) conducted a study in which a
multiracial group of women responded to a series of items drawn from a standard self-report measure of sex role stereotypes. Afterwards, they were also asked to indicate which of several possible definitions of key words in the items they had in mind when they answered the question. As a whole, there was no difference in how white women and women of color answered the original questions. However, for several items there were differences between the white women and the women of color and between women of different ethnic groups in what definition of the question they reported using when they answered. For example, the item “I am assertive” was sometimes defined as “saying whatever is on my mind,” “standing up for myself,” “expressing myself well,” or “aggressive.” Women of color were more likely to define being “assertive” as “saying whatever is on my mind” and white women were more likely to mean “standing up for myself” or “expressing myself well.” This kind of study, exploring what is meant by questions and answers, should be done more often. The original result—the finding of no group difference—is true because all of the specific definitions do correspond to accepted meanings of assertiveness. However, important nuances of meaning in responses can be overlooked even when face validity seems clear.

The concept of face validity is applicable both to individual questions used to collect data and to scales and measures composed of multiple items taken as a whole. When dealing with educational testing or assessment in particular, for example, the term content validity is also used. This concept deals with the face validity of a whole scale or measure and with the idea that its constituent items taken together adequately sample or cover all of the content relevant to the area being assessed. Consider the tests required to obtain a driver’s license: an eye examination, a written test consisting of questions about the state’s rules of the road, and a road test in which observations are made of the person driving in real world conditions and in which certain specific maneuvers, such as backing up and parallel parking, must be performed. The test has many parts specifically in order to assure that it has content validity, that all of the areas relevant to safe and competent driving are covered.

Both face validity and content validity address only the correspondence between data and concept through examination of the literal content of the data obtained. Another way to approach the issue of validity, however, is through examining multiple indicators of the same concept when they are available. The ways of evaluating validity that proceed in this way are termed criterion-related validity. A question or scale is said to have criterion-related validity when answers or scores derived from one question or scale are compared to other observations that have already been conceded to measure the concept in question. Clearly, as traditionally articulated this kind of assessment of validity is relevant only to fixed method research in which concepts can be predefined and measured in predetermined ways. Triangulation can be seen as a method used in qualitative research to address similar validity issues.

There are several forms of criterion-related validity. Establishing concurrent validity means that the answers to a question or set of questions are compared with a separate, often behavioral or real world measure of the same concept that the questions are designed to tap at the same point in time. For example, people
might both fill out a questionnaire designed to measure depression and be inter-
viewed and assessed by a mental health professional on the same visit to a clinic. 
The assessment of depression made by the mental health professional might be 
used to validate the depression questionnaire. If there were generally a high rate 
of agreement between the questionnaire results and the professional assessment 
(the criterion), this would be taken as evidence of concurrent validity. Predictive 
validity proceeds in the same way as concurrent validity except that the real life or 
other criterion does not occur until some time after the original data have been 
gathered. The questions are answered at one point in time; the criterion they are 
designed to measure does not occur until later. However, the logic is the same: 
The closer the correspondence of the answers to the questions with the criterion 
they are designed to predict, the higher the predictive validity is said to be.

Criterion-related validity, both concurrent and predictive, is most relevant when 
the phenomenon of interest can be tied clearly and unequivocally to a single, specific 
criterion, generally a behavioral one. Examples of such instances are the relationship 
of the SAT scores of college-bound high school students to the grade point averages 
earned in college or of a measure of perceived health risk to smoking behavior. 
However, many of the phenomena of interest in social work and the other helping 
professions are complex, socially determined, and difficult to relate to any one clear 
criterion in the present or in the future. In fact, many of the concepts of interest are by 
definition not directly observable. Consider “aggression,” for example: A comprehen-
sive definition of that phenomenon would likely include a consideration of action, 
context, and motivation or intent. As with the study of answers to a question about 
assertion discussed above, it can often be difficult to be sure that a specific behavior 
observed unequivocally is (or is not) evidence of aggression.

When considering complex phenomena that derive their meaning as much 
from theory as from any one piece of data, validity may need to be considered in 
relation to several other complex indicators, none of which may provide a perfect 
match conceptually or behaviorally with the phenomenon being measured. This 
form of establishing validity through comparing multiple measures with similar 
meanings is termed construct validity. It consists of comparing findings based on 
one imperfect measure of a phenomenon with others, both those that are consid-
ered much like what was originally measured (to show convergent validity) and 
those that measure things that are somewhat different from it (to show discrimi-
nant validity). Negative case analysis as used in qualitative data collection and 
analysis can be seen as a procedure that addresses issues of discriminant validity.

These concepts are very important in fields such as psychometrics in which 
the effort is to develop measures of psychological phenomena that can be used in 
fixed method research (see chapter 15 on selecting existing measures). Often 
evidence from measures using different methods of data collection—self-report 
compared with observation, for example—is used in establishing the construct 
validity of a measure in this way. Evidence for the construct validity of an existing 
instrument accumulates gradually based on reports of all the studies in which it 
has been used and compared to other indicators.

In the end, all the methods of assessing the validity of a data collection proce-
dure are dedicated to determining what the data collected in a certain way mean,
what theoretical concept or phenomenon they most accurately represent, what name the phenomenon observed should best be called by. In fixed method research, where methods of data collection can be standardized and repeated, a variety of ways of assessing validity—some data-based and some not—have been described. In flexible method research, the assessment of validity can occur both in the data gathering and in the analysis process. In the data gathering process in participant observation research, for example, the total experience that the researcher has of asking questions, listening, and observing helps to determine what the data mean. An answer to a question delivered with a smirk or without much feeling can be detected and a new approach to the topic devised until the researcher is convinced that the information offered is trustworthy and truly relevant to the topic at hand.

The process of data analysis, discussed at length in chapter 16, is as central to establishing the validity of data in flexible method research as is the process of gathering the data to begin with. In flexible method research, the processes of data collection and analysis or interpretation are often ongoing simultaneously; for example, a statement in the data found on reflection to be ambiguous can sometimes be explored and clarified on a subsequent occasion of data collection. Strauss (1987) speaks of an “indicator-concept model” of analysis in which multiple indicators or pieces of data are connected to a single given concept through coding. The names of the codes or categories of meaning used to organize the data may be in vivo, that is, supplied by one or more participants, or they may be constructed, that is, invented by the coder (Strauss 1987). In flexible method research, then, validity is a product not only of the data collection process but of the analysis process as well because that is when the concepts used to define the phenomena observed are designated.

**Summary**

Data collection is at the heart of all research, and there are many things to consider when planning, conducting, or evaluating any data collection effort. This chapter has covered the dimensions of decision making about data collection: the what, where, when, who, and how. Most important, of course, is determining what data will be collected, a decision that is determined by the purpose and focus of the research—its design, guiding question, and conceptual underpinning.

In particular, concepts of reliability and validity and the trustworthiness of data have been emphasized. These ideas are both simple and profound. Are the observations made and reported the products of the accidental or specific circumstances of the data collection or do they reflect some more essential features of the phenomena being studied? Are the phenomena that are described called by the right name, that is, are the observations connected with concepts in a way that seems defensible? What do the data actually represent? Data collection of any kind must be designed and carried out in a way that results in the gathering of information that is trustworthy, both reliable and valid. The next set of chapters in the book discuss in depth the three most common methods for collecting original data in social and psychological research—observation, interviewing, and questionnaires—in light of the basic principles of data collection outlined here.
Part IV
Methods of Data Collection

This section of the book discusses in depth the three main methods of collecting original data in social work and human services research—observation, interviewing, and using questionnaires and other forms of written self-reports. These chapters are written from the point of view of someone who is planning to do research and thus to develop methods and tools for data collection. The final chapter in the section discusses how to find, select among, and use existing measures instead. Chapter 11 in the previous section of the book, “Basic Issues in Data Collection,” presents the framework and key terminology used in these chapters. It should be consulted prior to using any of the material in this section of the text.
12
Observation

In some ways, all research rests on observation. Research is after all always empirical, and empirical by definition means perceived through the senses, that is, observed. In research, however, the term observation has come to mean something more restricted; it refers to data collected by direct visual experience of the study participant(s) by the researcher. Observation as a data collection method is underutilized in social work and human services research. Nevertheless, it is a vital tool for research that has been widely used in psychology and in such theoretically crucial areas as infant and child development. It has also been used to study social processes in formal organizations, informal groups, and even households.

In most cases, the person whose characteristic(s) are observed and the person doing the observing are different people. For this reason, observation as a data collection method is often assumed to be less biased than other methods of data collection based on self-reports. Stern (1985), for example, has commented that we live in an age in which observations are often the preferred form of evidence. As will become clear later on, however, even with observers who have no apparent reason to have a personal interest in the outcomes of their observations, expectations, and other characteristics of the observer can exert a powerful influence on what is thought to be “seen.”

There is something quite compelling about observational data. Critics may object to the manner in which observational data were collected or the interpretations that were made of them, but even the most skeptical critic of a researcher’s work is unlikely to discount totally data that were seen. In an enterprise where evidence is everything, observational evidence is often considered the most valuable of all.

None of the social sciences can lay claim to special expertise in observation as a data collection method. This measurement form is as old as science itself. What the social sciences can lay claim to, however, is refinement of this data collection method for assessing psychologically and socially defined abstract phenomena.
This refinement has been no easy task. The social sciences typically study phenomena that do not exist apart from conceptual and social judgments about them. This statement is especially true of the helping professions, which tend to study things when there is a socially defined need to do so.

Consider the act of striking a child. Once considered proper discipline in school and at home, this act is now considered abusive. It is relatively easy for someone who is present at the time to decide whether or not a child is being hit. It may be less easy to decide whether or not a bruise or other mark observed on a child’s body is evidence of having been hit or of some other kind of event, and it is even more difficult to decide what is good discipline in general and what is not. Even legal definitions of child abuse may be difficult to tie unambiguously to observational data. In social work, definitions of key phenomena can change over time and in different contexts, and the kind of data needed to provide irrefutable evidence even of an often-defined concept like child abuse may be hard to provide unambiguously. For example, Southeast Asian American women may use “cupping” and burning as folk cures for colds, remedies and acts of caring that are standard and accepted in their cultures. The marks these practices may leave on the body of the child may be seen as evidence of abuse although the same evidence would be labeled by others in their own communities as evidence of nurturance (Landrine, Klonoff, & Brown-Collins 1992).

This chapter gives an overview of observation as a data collection method as it used in both flexible and fixed methods of research. In flexible method research the observations made are formative of the understandings and explanations generated. In fixed method research, by contrast, the phenomena to be observed are specified and defined before data collection begins, and the observational process is structured in light of those specifications and definitions. In fact, the observational traditions in fixed method research and in flexible method research are based on sharply divergent assumptions about research, the nature of useful evidence, and the role of the researcher in observation. What observation has in common in both design contexts, however, is its focus on nonverbal, visual data and the dilemmas that are raised by the observation process itself.

**Basic Principles of Observation**

In the helping professions, observational methodologies have largely been developed and refined by clinicians and researchers working within the behavioral tradition. This connection is not surprising: Behavioral theorists traditionally held that practitioners should address what can be seen, that is, overt behavior. As a consequence, their measurement methods, largely directed toward assessing the observable, have been predominantly observational ones. Behavioral research has generally followed traditional models and used fixed methods of research.

In traditional behavioral terms, observation is a process by which an instrument is used to calibrate certain predefined properties of an observed and distinct other. The instrument of observation may be mechanical, such as a machine to measure galvanic skin response, reflecting perspiration as an index of physiological anxiety.
Actually, the term *detection* rather than *observation* should be used when instruments function to record phenomena that would ordinarily be imperceptible.

More often than not, however, the instrument of observation is a person, and any instrument used, such as a videotape recorder, functions only to make a record of what human senses would perceive. However, when people function as observers in the behavioral tradition, they are instructed do so as precisely and objectively as a machine. That is, the human observer makes a record of what is observed by rules that must be invariantly applied, and he or she also must minimize the impact on or interaction with the observed above and beyond what may be inherent in the act of observation itself.

To some extent, however, observation involves making judgments. Traditionally, observers are therefore trained in applying some clear rules to their observations, rules that specify:

1. The circumstances of observation, when and where observations will take place, and how much control will be exercised over the observational context;
2. The period of time over which observations will be made and with what frequency, including whether and how to sample times of observation;
3. Precisely who will be observed and what will be observed about each person; and
4. How the observations will be recorded.

Even when observations are less structured, decisions are made about each of these dimensions. Participant observational research, however, is opportunistic (Jorgensen 1989), meaning that these decisions are often made in context in the field and adjusted as the process unfolds. In fixed method research, decisions about these dimensions are made before data collection begins. Whatever the form of the research, however, the guiding principle used when answering these questions, of course, is the purpose of the research prompting the observation.

*Context and Circumstance*

With respect to the question of when observations should be made, the optimal circumstances of observation in flexible method research are whatever ones present themselves in the setting that are relevant to the study question(s). The principle of prolonged engagement (Padgett 1998) indicates that observations are often made over an extended period of time. In fixed method research, the question of when to observe is essentially a matter of deciding on the context(s) in which the phenomena of interest can best and most efficiently be assessed.

The question of how long to observe really breaks down into two different questions, depending on whether the researcher intends to observe a few people over an extended period of time (as is often the case in flexible method research and single case designs) or many people at one or only a few points in time. In the first case, the question remains how long these few people should be observed. In the second the question becomes how many people to observe; since each person will be observed for a roughly equivalent amount of time, the total amount of observation time will be determined by the number of different people observed.
In flexible method research, it is not possible to specify for certain how much observation will be enough before the fact. Leaving such matters open is inherent in this form of design, since the method must remain flexible enough to respond to the data as they come in. With flexible method research, then, the answer to the question of how long to collect observations is at the same time simple and demanding: Until enough data have been gathered to answer the questions to be addressed with convincing data to defend the answer, or at least until additional observation time is yielding little or no new data or insights. That is why how much observation will be enough cannot be specified exactly beforehand.

This inability to specify before the fact how much observation will be enough, which is inherent in flexible method research, is also inherent in single-subject studies (see chapter 8). As will be recalled, single-subject designs set out to document specific changes in behavior that occur following an intervention. In these designs, as in flexible method research, the length of time over which data must be collected is determined not a priori, but rather by the form the data are taking as they occur: Data collection in single-subject designs must continue until the response to the intervention becomes clear. And while sometimes there is an immediate and dramatic reaction to an intervention, more often than not data collection must continue for some period of time until any trends or changes in behavior during the intervention period can be adequately documented. In both single-subject designs and flexible method designs, then, the answer to the question “How much should I observe” is the same: you should observe until you have an answer to the question or hypothesis you set out to address, and until you have data in hand that you can present to others that will give them reason to believe you and agree with your conclusions.

The question of how long to observe can be answered before the fact more easily in descriptive, relational, and group experimental designs. The question really is a conceptual one that boils down to whether or not the phenomena being studied require a longitudinal approach to answer the question or whether a cross-sectional design can serve the purpose (see chapter 6).

Time is only one element of the observational context. The circumstances or settings in which the observations will take place must be considered as well. In any type of research, then, the researcher must decide before initiating the observations what the observational circumstances will be.

Whether planning a fixed or flexible method study, preliminary work is often done to determine which of several alternative available observational circumstances may present the best ones for the research. In field research, this work typically involves finding naturally occurring settings where the behavior of interest is likely to occur and that will accommodate the observation process. In fixed method research, the goal may also be to identify or arrange a setting that will remain sufficiently stable and available to allow making meaningful comparisons of behavior across repeated observations made in it.

Since fixed method research is often directed toward making some form of comparison, it frequently requires repeated observations either over time or across people. Comparing what is observed in these repeated observations in a way that permits drawing conclusions about the people who were observed rather than the
Observation

• 331
circumstances of observation means that influential elements of the observational context must remain the same or at least be equivalent. If constancy or at least equivalence of the observational context is not maintained, what is observed can be expected to literally look different, not because it is in fact different in some intrinsic way, but rather because the circumstances under which it is being studied have changed.

Since comparability of observational setting can be so important to fixed method researchers, sometimes the researcher decides to create the observational circumstance rather than using a naturally occurring one. Consistency in the observational context can be arranged either within or outside the laboratory. When it is arranged within the laboratory, the form of the observation is typically called controlled, because the researcher is controlling not what the person being observed is doing but the circumstances within which the participant can do whatever he or she is doing. When it is arranged outside the laboratory, it is called contrived. There is no functional difference between controlled and contrived observation: Both involve standardizing the observational context. The difference between them is that one takes place within a laboratory, and therefore always with the knowledge of the person being observed, while the other takes place outside the laboratory, and therefore potentially without the knowledge of the person being observed.

One example of this use of a controlled observational context is in attachment research (Ainsworth, Blehar, Waters, & Wall 1978). Their strange-situation procedure was carefully designed to assess the attachment of one-year old babies to their primary caretaker, the mother. Using a one-way mirror observation is made of the mother. Two chairs for adults, one for the child, and some toys are arranged in a standardized way in an office-like room. A fixed sequence of events takes place, beginning with the mother and baby alone in the room, the entry of a stranger who approaches the baby and then leaves, the leaving of the mother, the return of the stranger, and the final return of the mother. Babies react in quite different ways to this sequence of events. Their contact with the adults, their movement around the room, and their affect are all observed by two people who are not in the room. A system of categories has been developed to describe the various styles of reactions that babies have displayed in this standardized set of circumstances. This classification system for styles of attachment in infancy, in turn, has subsequently been studied as a predictor of many aspects of later development. This is only one example of how controlled observational circumstances have been used in the study of infant and child behavior and development.

Whom and What to Observe

Whom to observe would at first seem obvious: Certainly the people chosen to be participants in the research should be observed. As will be recalled from the sampling chapter, these participants should be ones in whom the research phenomenon is expected to appear in clear, even exaggerated form or people representative of the larger group to which one hopes to generalize.

Given this, the question becomes “Who else to observe?” When the research question involves inquiry into the context surrounding the primary participant’s behavior, it is often appropriate to observe people other than the primary
participant as well. From this perspective, these other people become important as potentially interesting features of the interpersonal context for the primary participant’s observed behavior. These other people become part of the “what” that is observed, the Jarret (1992) study in chapter 4 illustrates.

The behavioral tradition has given rise to a form of observation in which the observer and the observed are one and the same. This form of direct observation is called self-monitoring or self-observation. Typically self-monitoring is used with one of two types of behaviors: Ones that are easily defined and counted, such as the number of cigarettes smoked each day, and subjective experiences that are otherwise inaccessible to outside observers, such as the number of self-critical thoughts while smoking a cigarette. This chapter, however, focuses on observation by another and thus on how to capture behavior and interactions, those things that are observable by someone else.

Of the several questions that must be answered when designing or choosing an observational methodology, it is the what question that has received most careful attention, but this attention has tended to be technical. Much has been written about characteristics that distinguish between phenomena that can be observed successfully and those that cannot. In general, physical objects, nonverbal behaviors, facial expressions, gestures, and social interactions lend themselves best to observation. On the other hand, subjective experiences, ideas, meanings and other intangibles can only be inferred from observational data, and other modes of data collection may be preferable for studying such phenomena. However, children and others who cannot give a verbal account of themselves understandable to the observer cannot be interviewed, and observation is often used to study them instead. In addition, what people do and what they know or say they do may differ markedly. Much social and interactive behavior may be unconscious and unexamined by the participants, and observational methods can be more effective for studying them than relying on some form of self-report.

A major challenge in observing an abstract phenomenon is developing a definition for the phenomenon that discriminates between when that phenomenon as defined is present and when it is not. In fixed method research, it is assumed that the phenomenon to be observed has been adequately defined, at least at a conceptual level. It should be remembered, however, that the nature of the definition critically and indelibly influences the nature of the data subsequently collected. As is always the case in data collection, the what that is being measured can never be assumed to be correct only on the basis of the “how.”

Observational data describing well-defined, observable phenomena, however, only seem self-evident; the complexity of the act of observation has traditionally been overlooked. The logical positivist tradition in social science research maintained that observers could see what they were observing in an objective and undistorted fashion. Moreover, they assumed that when they made observations, what they could see was raw reality, so that what they believed they saw would reflect what was objectively really there.

This belief reflected two underlying assumptions. The first was that human observers could be made to function as reliably and predictably as would a mechanical recorder: Human observers could be trained well enough so that they
would perform just like a reliable thermometer, which if used repeatedly in the same fashion—located correctly and left in place for the proper amount of time—would calibrate the temperature of a feverish person with the same result given the same level of fever each time.¹ This first assumption turns out to be true only sometimes. In fixed method research, it has been shown that human observers are reliable only when rigorous design guidelines are followed when both developing and applying whatever observational methodology is being used.

There is a second assumption embedded in the logical positivist tradition, however, which turns out to be much more problematic: Human observers can be passive receptors of their external worlds. Observations are only what the observer makes of them. Stern (1985), for example, speaks of the “observed infant” described in developmental psychology as “a special construct, a description of capacities that can be observed directly” (p. 17). Such a view of the infant is formed both by what the observer notices and by what the observer can notice; the infant’s subjective experience, for example, can only be inferred.

The way in which observations are processed influences what is seen. This effect is more subtle, but its impact is more profound. For a moment, force your attention away from the words you are reading in this text. Without changing positions, let your eyes move around the space in front of you. No matter where you are sitting, there will be a great number of objects and surfaces before you. Had you noticed them when you were reading the text? Probably not. The point is that when we observe, we engage in a process similar to the one you were using earlier when reading this text: We see only a portion of what is in front of us, the parts we choose to focus on.

It is easy to see how this form of perceptual selection bias could exist when evaluating a very focussed, specific observational system. In fact, there is some reason to believe that the more focussed an observational system is, the less possible it is to notice phenomena that were “unexpected.” By defining what is to be observed, constraints are placed on what can be observed. Observers are unlikely to see what they are not looking for.

Decisions about what to observe, then, are of profound importance. They must be made so that the observational system chosen will be consistent in principle with the properties of the research design. In flexible method research, observational systems must start out as very unfocussed. The observations are always guided by a general notion of what is being looked for, but to have specific notions of what will form the phenomena of interest would likely preclude the observer’s being sensitive to the unexpected. And since flexible method research has as its intent understanding things that are poorly or incompletely understood, which implies seeing the unexpected, to use an observational system with preconceived parameters would not work. Mutual development of method and data thus characterizes the observational process used in flexible method research.

¹. This is, of course, an oversimplification. As was emphasized in chapter 11, there is some error inherent in all measurements, including observations. It would be more technically accurate to note that thermometers, when they are functioning reliably, give approximately the same temperature readings given the same level of fever each time they are used.
At this point it should come as no surprise that it may be easier to conduct observations within the frame of descriptive, relational, and experimental research than in a more open-ended fashion. These three research designs all require specifying fixed methods, methods framed around specific research questions or hypotheses, before the data are in hand. It follows, then, that the specific variables that will be observed not only can but should be specified and defined beforehand.

Recording Observations

The question of how observations should be recorded is, of course, a technical one with a general answer that applies in all cases, regardless of the type of research or the observational methodology employed. The method that preserves as much of the available relevant data as possible with the greatest degree of accuracy and permanency and the least degree of intrusion on the data preserved is the one that should be used. Since most observations are collected and recorded by human observers, we have grown accustomed to collapsing the several stages in an observe-record-code complex into one in our thinking, but they are really separate stages, each of which merits attention.

Consider the situation in which observation is accomplished by a video camera. When the camera is running, observation is taking place in the situation, even if there is a mechanical problem precluding making a record of what is observed. If there is no mechanical failure, there will be a record on the tape. However, what will be recorded is limited to the times when the camera was running and the part of the scene on which the camera was focussed. In addition, not until those data are coded, or translated into meaning units in some fashion, is the observation of them complete. It clarifies the issues involved in the observation-record-code complex to consider each of the facets in it separately.

The recording process refers to the technology used to make a permanent record of what is observed. This technology is most clearly illustrated by the photographic methodology of video cameras when video observation is used or the magnetic recording processes of audio recorders when audio observation is used. Often, however, the recording methodology involves arranging for a human being to serve in the role of an audio and/or video recorder by providing that human observer with a set of rules indicating what record to make when a certain event is observed to occur.

There is a clear analogue between the mechanical and human processes in observation. Mechanical technology works by virtue of the translation of information from one form to another. In the case of audiotape recordings, for example, sounds issued within a certain frequency range and greater than a certain volume will be recorded using a machine that enters electrical impulses onto magnetic tape, which then becomes a symbolic representation of the event transpired. In like fashion, when human observers make records of what is observed, they follow a set of rules that results in their entering symbols usually onto paper such that these symbols are recognized as representing the event recorded.

In some circumstances, researchers prefer using mechanical rather than human observers. With human observers, there are lapses in attention or motiva-
Observation and misunderstandings or differences of interpretation about how to implement the recording rules. In addition, human recorders have been shown to change over time: As they observe a certain phenomenon more often, they can become immune to noticing minor instances of it, which has been termed “calibration shift” (Kent & Foster 1977).

Unlike humans, mechanical recorders are consistent. However, they lack some capacities that make human recorders an indispensable part often of the data recording and always of the data reduction process. These capacities include selectivity in recording what is observed, sensitivity to what was not expected to be observed, flexibility in changing focus when needed, and the capacity for following recording rules that involve higher order processing of information.

Machines record indiscriminately. For example, a machine will, and can only, record an entire conversation, while a human observer can be instructed and trained to record only those parts of a conversation characterized by a certain content, for example, those parts relevant to the expression of affection. Humans, are thinking recorders, and having a judging being making records can be an asset in observation.

Research questions often concern complex phenomena that are difficult to define simply. Quite often determining whether the researched phenomenon has occurred requires making complex judgments. Machines will only record what they are sensitive to recording. An audiotape, for example, will miss whatever nonverbal communication is taking place, and a video camera will record only what is focussed on and only for the period specified. Human recorders, on the other hand, do not lose their more general awareness when following a set of recording rules. Unexpected or complex phenomena that strike the human recorder as important can be noted, observed, and passed on to the investigator.

The process of coding is always a human activity. It involves selecting specific features of the observation record to classify conceptually. Observational systems that rely on human observers typically complete the record and encode components of the observation-record-encode complex simultaneously; in fact, typically only codable features of what is observed are recorded as part of the observation process. Observational systems that rely on mechanical recorders, on the other hand, accomplish the recording and encoding components separately. The former is handled mechanically while the latter is managed by a person after the fact.

The Observer and the Observed

There is a final, critical feature of observation that must be addressed: the relationship between the observer and the observed. In the traditional view, observation implied that there was no relationship, that the observer was related to the observed only through passively receiving and accurately recording information about the behavior of those observed. There are of course very few observational circumstances that approximate this situation.

Consider what can happen when making observations of two children interacting with one of their parents. The observer’s role involves sitting quietly in a
corner, clipboard and pencil in hand, observing and recording any event that seemed important in understanding the parent-child interaction. The children had a different take on the circumstance. The observer was not to be allowed to remain in the role of passive, uninvolved environmental feature. Questions ranging from “Who are you?” to “What are you writing?” are asked continuously. When the questions go unanswered, the observer is subjected to nose pulls, shin kicks, and lap climbs. Under such circumstances, an observer will usually find it impossible to maintain a totally passive stance.

Reactivity

In general, the process of observation involves introducing something into a situation that would not otherwise be there. As has been mentioned many times, the introduction of this feature constitutes making a change in the circumstances surrounding the researched phenomenon. And since we are typically interested in studying phenomena in people who are responsive to their environments, the process of observation can be expected to affect the person being observed and often the feature(s) of interest in the person being observed. This effect is termed reactivity; it refers to change induced in what is observed by the process of observing it.

The reactivity inherent in observation troubles many researchers: No one likes to think that he or she is getting an invalid picture of what is being studied. It is distress over this nettling reality that has led some researchers to strive for concealed means of observation. An example of such concealed observation is presented by the work of Humphreys (1970), who entered public restroom facilities frequented by men who came there to engage in impersonal acts of homosexual sex to systematically gather data on their activities. People observed in the rest room were traced by means of their auto license plate numbers and found to include middle class, married members of the community. The use of a concealed observational method was defended on the basis of reactivity: Given the nature of the behavior being observed, the phenomenon studied would never have been evidenced in the presence of a known observer. Humphreys took on a role in the setting—of lookout—that justified his presence as an otherwise nonparticipating observer, a strategy he defended as “passing as deviant to avoid disrupting the behavior he wished to observe” but not for the purpose of gaining access to a space that was private, since he was already entitled to access (Golden 1976).

This study presents a dramatic and controversial example highlighting the troubling ethical issues raised by the question of whether researchers ever have the right to observe others without their knowledge and informed consent. Most Institutional Review Boards believe that such observation is permissible only when studying innocuous behaviors as they naturally occur in very public situations and when the persons observed are not identified or affected by the observers. Concealed observation in other cases is more controversial and represents a clear example of the important role Institutional Review Boards, or consultation with community members and colleagues not invested in the research plan, have to play in monitoring research activities.
The observer-observed relationship distills into a single question: How much and what kinds of interactions will occur within it? Some observational systems try to minimize or standardize them. In these systems, the observer is to function in a manner analogous to a mechanical instrument. As much as possible, the observer is to be affected by but not to affect the observed. Any effect is to be only temporary and is to extend only as far as is necessary to make a record of whatever observed event was just noticed. However, other observational systems allow for, expect, and in fact make use of observer-observed interaction to further the research. One such form of observation is called participant observation, which is discussed in more detail below.

Examples of Observation in Use

In recent decades the use of observation, generally of a quite structured kind, has revolutionized the field of early infant research and the theoretical understanding of psychological development in infancy. Fraiberg’s (1970) research on blind infants illustrates the use of observation, structured and unstructured, in developmental research. Stern (1985) describes the new developments in infant research as follows:

The revolution in research consisted of turning the situation on its head, by asking not, what is a good question to pose to an infant? But what might an infant be able to do (like sucking) that would serve as an answer? With this simple turnaround, the search for infant abilities that could be made into answers (response measures) began, and the revolution was set in motion. . . . Good infant “answers” have to be readily observable behaviors that are frequently performed, that are under voluntary muscular control, and that can be solicited during alert inactivity. Three such behavioral answers immediately qualify, beginning at birth: head-turning, sucking and looking (pp. 38–39).

Infants, for example, have been shown, using electronically wired pacifiers, to be especially interested in the human voice as compared to other sounds of the same pitch and loudness. They also demonstrate by turning their heads that they prefer the smell of their own mother’s milk to the milk of another woman (Stern 1985). From the accumulation of many such observational studies of infants’ early abilities, preferences, and interests, new theoretical understandings of infant development and parent-child interactions have emerged.

Infant research is not the only area in which observational data are useful, however. Jarrett’s (1992) case study, reprinted in chapter 4, gave one example of observation in use. As another example, Hochschild’s (1989) study of how two-career couples handle the work of caring for a home and young children used in-home observation to supplement the extensive data collected through interviewing. The purpose of Hochschild’s study was to explore whether or not married men were contributing any more to household work than the early studies in the 1960s and 1970s had shown. She also explored the various tensions that
were occurring in couples because of class norms, ideologies about how things “should” be that didn’t match with day-to-day necessity, and the effects of earnings differentials in couples in how work at home was shared.

In particular, she wished to examine how well people’s accounts of themselves and how they shared domestic responsibilities meshed with reality. In part this was accomplished by interviewing both members of a couple, but observation of selected couples at home was done as well:

I also watched daily life in a dozen homes during a weekday evening, during the weekend, and during the months that followed, when I was invited on outings, to dinner, or just to talk. I found myself waiting on the front doorstep as weary parents and hungry children tumbled out of the family car. . . . I sat on the living-room floor and drew pictures and played with the children. I watched as parents gave them baths, read stories, and said goodnight. Most couples tried to bring me in to the family scene, inviting me to eat with them and talk. I responded if they spoke to me, from time to time asked questions, but I rarely initiated conversations. I tried to become as unobtrusive as the family dog. . . (Hochschild 1989:6–7).

Note that this brief description offers a sense of the “time sampling” attempted with each family. It also speaks to the particular role that she as observer elected to play in each setting.

What did such an observational strategy yield? In many cases, subtle effects could be seen of the strains and tensions between husbands and wives about housework (and other issues) that they might not even be aware of. A brief vignette from one family may illustrate:

After a long day, mother, father, and son sit down to dinner. Evan and Nancy get the first chance of the day to talk to each other, but both turn anxiously to Joey, expecting his mood to deteriorate. Nancy asks him if he wants celery with peanut butter on it. Joey says yes. “Are you sure that’s how you want it?” “Yes.” Then the fidgeting begins. “I don’t like the strings on my celery.” “Celery is made up of strings.” “The celery is too big.” Nancy grimly slices the celery. A certain tension mounts. Every time one parent begins a conversation with the other, Joey interrupts. “I don’t have anything to drink.” . . . By the end of the meal, no one has obstructed Joey’s victory . . . (Hochschild 1989:35).

If nothing else, such observation illustrates vividly the vicissitudes of daily life in the family. Hochschild uses such data, however, to make the point that inequities in the division of child care and housework responsibilities set up complicated dynamics among all family members that help to perpetuate whatever arrangement exists. Based on the interview data, the mother and father in this family had a wholly different explanation for “Joey’s problem” that he wouldn’t go to bed at night, as “normal,” stage-related behavior. Whatever the framework that might be used to explain the interactions observed, the observational data make the nature of them quite vivid and clearly suggest, based on an interpretation of the observer,
that Joey’s behavior is related to dynamics between his mother and father. Thus observational data can reveal aspects of a situation that even the participants in it may not see or articulate clearly.

Developing or Selecting a Structured Observational Tool

As is true with all data collection methods, a researcher has two options when planning observational research in the context of a fixed method study. He or she may choose to employ a measure, in this case an observational system, previously developed by another investigator, or he or she may choose to develop a new observational system specifically tailored to the requirements of his or her research circumstance. Chapter 15 covers how to locate and select an existing measure. Naturally, whenever the research design is a flexible one, the form of the research requires developing the observational system as the research goes along. Hochschild (1989), for example, found that talking to a family’s babysitters and child care providers and even doing some observation with them was an unexpected and useful source of information, and she then added this to her plan of data gathering for each family. A skilled researcher, then, should know how to invent an observational methodology. Considering how a new observational system has to be developed will clarify many of the issues that should be attended to when deciding whether someone else’s system is good enough to use.

Developing an Observational System

The first step in observation is specifying the purpose of the observation, what question(s) the observation is supposed to inform. In flexible method research, this purpose will usually be to define and understand the phenomena of interest more clearly. In other forms of research, this purpose will be to describe and/or show connections between some already well-defined research phenomena. In the latter cases, this first step implies clearly specifying the definition of the phenomenon to be researched or observed. In the beginning, this definition properly remains an abstract, conceptual one. The definition will be narrowed to an operational one, that is, to a set of rules for deciding whether or not or how much of a phenomenon has been seen in response to what is observed.

It may be of interest in this context to note that developing a new observational system for any research purpose often requires conducting what might most easily be described as flexible method research at the outset. While there is a general idea of what is to be observed, an idea that is used to guide the observation process, specific notions of what is to be noticed are left open until the data are given an opportunity to inform and shape them. The Fraiberg (1970) studies of the development of infants blind from birth illustrate this evolution well when she uses her observations of a particular case to shape her later observations of a group of blind babies.

Any research study takes place in a specific context. That is, the question or hypothesis will be posed not in the abstract but under some narrowed or restricted
set of circumstances. These circumstances must also be defined at the outset of developing an observational system. Just as the research purpose guides what is to be noticed, the context defines the circumstances under which observations will be made.

Practical considerations enter into the selection of an observational context as well. Suppose one is interested in studying intimacy. Optimal circumstances for making observations relevant to this construct would most probably be private (i.e., intimate) ones. However, short of developing a mechanical observation system that can run continuously in such circumstances (and this procedure has in fact been used), arranging observations in private circumstances is quite difficult and can raise important ethical questions.

Once the researcher has specified the purpose of the observation, which in the case of descriptive, relational, and experimental work will include stating an explicit conceptual definition of the phenomenon to be observed and the circumstances under which observations will be made, unstructured observations should take place. In unstructured observations the researcher enters the observational situation armed only with a clipboard and a pen or pencil or a portable computer and the guiding notion of the research purpose. The observer’s initial task is to literally observe, to notice everything taking place in the circumstance that seems even remotely related to the research purpose as stated, and to record as completely as possible a narrative description of what he or she sees.

The observer’s task does not involve interpreting or making sense of what is noticed at this point. As much as possible, the observer is to function as a probe, an instrument of observation sufficiently sensitive to notice whatever important is there to be observed. Preconceived notions and theoretical biases should exert as minimal an influence as possible on what is noticed.

Although there is no such thing as a “naïve observer,” since without some sophistication an observer would be incapable of discriminating between what information it is important to attend to and what information is not, observers will necessarily have different perspectives. In some cases, these perspectives can be dictated strongly by preconceived notions. To the extent that they are, they become biases, that is, predispositions to “see” events in a certain way, even when the events are seen in a different way by others with a different perspective. It can even be useful to employ multiple observers selected so as to vary importantly from one another. When research is done by a team, it is always a good idea for the person directing the research, who may have the strongest preconceptions about what should or will be seen, to conduct some of this informal observation as well.

In fixed-method research, the data obtained in this way are regarded not as an end in themselves but instead as preliminary. The records from these observations must be studied carefully to identify observations that are concrete instances of the research phenomena to be studied. The descriptions of these observed events are then used to derive statements identifying how to recognize instances of the researched phenomenon and how to distinguish between observed events that are instances of it and similar events that are not.

For example, suppose a fixed method study was aimed at observing “aggression.” A conceptual definition for this research phenomenon might be “actions
with the apparent primary intent of coercing another or expressing anger toward
them.” While this conceptual definition might be acceptable to a variety of
researchers studying aggression in a variety of contexts, concrete definitions of
aggression would be expected to vary widely across subject types and observa-
tional situations. What is aggressive behavior among four year olds, for example,
is quite different from what is aggressive behavior in adults. Similarly, the form
adult aggression may take at work typically differs from the form adult aggression
may take at home. Variations in concrete definitions might also arise because of
differences in theoretical or political perspectives: what appears to be aggression
from one person’s perspective might appear to be self-defense from another.
Sorting through narrative records from unstructured observations, then, allows a
fixed method researcher to clarify what the distinguishing features of the research
phenomenon are, given his or her interpretation of the conceptual definition and
given the people and setting that will be observed.

Once these distinguishing features are identified, the statements describing
them become the operational definitions of the research phenomenon. These defi-
nitions then form the basis of what observers will be trained and expected to
recognize. The purpose of the initial unstructured observations is to generate
information that will allow the researcher to create an informed system for struc-
tured observations. In this way, an observational system can be tailor-made to fit
the observational circumstances, making it more likely that the structured system
will be sensitive to the research phenomenon as it is conceptually defined and as it
is manifested in the particular circumstances that will be studied.

Once the definition has been clearly stated, the technology surrounding how
the defined phenomenon will be observed, recorded, and coded must be specified.
The observation technology will typically be either mechanical or human. Some-
times the phenomenon can be operationally defined such that a mechanical device
can reliably detect it, such as defining “making too much noise in the house” as
“sound episodes or events that exceed 70 decibels for three or more seconds.” In
such a case, if the appropriate mechanical apparatus is both available and accept-
able to those who will be observed, then a mechanical observational system is
often preferable.

Sometimes, however, the phenomenon as operationally defined is either too
subtle or too complex to be mechanically detected. This is particularly likely to be
the case in two circumstances. First, many psychological phenomena are difficult
to detect mechanically: these phenomena are typically quite subtle, and their indi-
cators can vary widely across different people or even within the same person at
different points in time. Second, many interpersonal phenomena can only be
detected by human observers, since recognition of the occurrence of such
phenomena typically cannot occur without a knowledge and understanding of the
interactional context within which particular events are occurring.

Where a mechanical observational system can be arranged, developing the
observe-record-code technology is quite straightforward. The apparatus is posi-
tioned optimally, that is, so as to have most complete access to the circumstances
of observation, and an observation schedule is set. The observation schedule can
be continuous, which is most appropriate when a research phenomenon both
rarely occurs and is no more likely to occur at certain times than others. If one were researching how two-worker families with school-age children manage morning tasks, however, it would be inefficient to record all day—continuously—rather than just during the morning hours when the research phenomenon is present. But if one were researching all the ways in which such parents coped with both work and family responsibilities, it would be optimal to record continuously at various times in the day except, of course, when both were at work, as Hochschild (1989) did.

**Sampling Observational Periods**

In general, with any circumstance where the research phenomenon appears often, observations can be conducted on a schedule rather than all the time; they can be sampled. The assumption under these circumstances, of course, is that the times that are observed give a representative picture of what happens during the times that are not observed. This assumption, it will be recalled, is similar to the assumption always made whenever a sample is used to draw conclusions about a larger population. In other words, study participants are not the only feature sampled in research designs; occasions for observation may be another. In this case, the assumption is that what is observed during those times that observation takes place represents fairly what would have been observed at other times that observation might or could have taken place.

As in sampling individuals, when the interest is in generalizing from the sample observations to some larger population of observations, the only or at least primary determinant of when observations occur should be chance. Optimally, then, one would decide on the duration of single observation sessions (for instance, 45 minutes), write each possible 45-minute observation period on a slip of paper, and then randomly select slips to determine when observations will be made each day.

Practical constraints typically impinge on this ideal. Especially with human observers, their schedules must be at least as influential a determinant of when observation sessions take place as is chance. And in certain settings, such as schools, hospitals, nursing homes, or homeless shelters, there are organizational schedules that dictate when observations can be made. It is not bad research when these practical considerations, rather than chance, determine when observations occur. But the conclusions drawn from research where observation schedules were determined by practical considerations must take into account the limits of these practical features on the generality of what was found.

Most fixed method observation systems or schedules specify short time periods during which the observer will work, followed by short time periods during which the observer will make a record of what was just seen. The length of both these intervals varies. The length of the observation cycle will depend on how much observation time is required to see enough of what is going on to be able to detect the research phenomenon as defined. This length of time will obviously need to be longer for phenomena that themselves take longer to occur, as well as for phenomena that cannot be identified without knowledge of what else is
going on when they do or do not occur. Someone observing compliance with
authority figures, for example, would be unable to code a particular action as an
instance of compliance without knowing whether or not the action had been
preceded by an authority figure request to do it. Recording cycles are usually
short. They need only be long enough to permit the observer to make a record of
whether the observation target events did or did not occur.

**Recording Data**

An observation record can be either mechanically or humanly produced. In the
case of mechanical observation, the record will of course be made mechanically.
A video recorder will produce a videotape record, an audio recorder will produce
an audiotape record, and a polygraph will produce a written polygram.

In the case of human observation, the record can be made either with or
without the assistance of mechanical aids. In the simplest case, the human
observer will keep track of what was observed by making marks on an observa-
tion sheet. A set of codes will appear for each observation period or cycle. At the
end of each cycle, the observer will mark various codes to reflect which of the
observed behaviors did and did not occur during each of the observation intervals.
If available, mechanical or computerized recording methods can be used to facili-
tate this process.

When the observations have been made by humans, the observe-record-code
complex has been completed with the creation of the record, assuming that the
recording system involves making a record by coding whether specific, predefined
phenomena were or were not observed during the observation cycle. Thus in
structured observation systems, to record *is* to code. When unstructured observa-
tions are made mechanically, however, the observation record, such as a video-
tape, will itself typically require review and coding by a human observer. In other
words, only the observe and record parts of the observe-record-code sequence
have occurred.

Having a videotape certainly provides certain research advantages. It is no
small contribution, for example, to have a permanent record of the events of
interest that can be examined and reexamined, permitting a greater degree of
depth in what is understood about them. Alongside this decided advantage is a
disadvantage. The videotape record is no more informative than were the events
as they originally occurred. Usually, then, mechanical observation requires a
human observer who must later execute the observe and code components of the
observe-record-code complex during a coding process.

Whether using a mechanically assisted observation technology or one that
relies only on human observers, the end product of the observe-record-code
process is information about whether the phenomena as defined did or did not
occur during each of a series of observation intervals. This information is typi-
cally translated into a frequency score, reflecting the number of times the research
phenomenon was seen during the observation period. As with any structured data
collection method, the resulting data are then analyzed for reliability and validity
as described below.
The advantages of developing a new observational system are clear: the system can be tailored to conform exactly to the new investigator’s conception of the researched phenomenon as well as the specific research circumstances involved. But this luxury comes at some cost: It takes a lot of effort to develop an observational system, as is the case with any new measure.

Using an Existing Observational Measure

Having discussed the steps that one must go through to create a good structured observational tool, it may seem obvious why fixed method research often uses existing measures. First, the quality of the observational system, in terms of its reliability and validity in other studies, will already have been established. It is generally easier to demonstrate a measurement tool’s usefulness in a new context than to demonstrate the reliability and validity of one that has never been used before.

There is a second reason for the use of preestablished systems whenever possible: Doing so allows comparisons of results across studies having measurement methods in common. Using a preestablished observational measure creates a connection between a new study and others that have used the same measure before. It is this interlocking of method, as well as of question, that makes research a collaborative rather than an isolated activity across investigators. Familiar methods yield results more easily interpretable by other researchers.

How to find and select among existing measures is covered in detail in chapter 15. In general, there is little difference between selecting an existing measure for observation as compared to an interview or questionnaire measure. However, there are a couple of additional points to consider. The first is that in addition to the content of the observations, the observe-record-code process must also be suited to the circumstances of the present investigation. The recording system in particular deserves review; some may be too demanding or complex for easy use in all settings. The training requirements for observers should also be considered. In general, when multiple observers are used, a rate of interobserver agreement of 85 percent is considered minimally acceptable.

As with other preexisting measures, the reliability and validity of any observational measure adopted for a study must be established in the new context of use and checked throughout the duration of the data gathering. In addition, it is also useful to arrange checks on what has traditionally been called “calibration slippage,” which refers to the fact that observers may unintentionally alter their standards of observation over time. This sort of slippage occurs, for example, when there is a certain event that strikes observers as a blatant example of what is being observed when the study first begins but as not a real example of what is being observed after they have watched things for a while, becoming “jaded” or “desensitized” to what they see.

Of course it is not always possible to use a previously developed observational system. Sometimes one is not available. Sometimes the ones available do not suit the circumstances, sample, or conceptual base of the research. In this latter circumstance, sometimes the existing system can be modified to make it acceptable: Observation intervals can be lengthened, recording technologies can
be added, and so forth. Whenever such changes are made, however, the observational method is in reality a new one, and its reliability and validity should be again examined and reported.

Reliability and Validity in Observation

It is interesting to note that, because observational data can seem so inherently credible, the reliability and validity of observational data collection systems have not always been examined as carefully as other types. There is increasing acknowledgement even among traditional measurement specialists that this failure is unfortunate: Every data collection method must be shown to be good and not assumed to be so.

The most usual form of reliability examined with observational data collection is called variously interrater reliability and interobserver agreement. This kind of reliability requires demonstrating that different raters or observers are making or would make the same recordings about the same events. Most often interrater reliability is demonstrated by arranging for two or more observers to make records of the same situations or events. In the case of live observations, this would involve being in the same place at the same time as Fraiberg (1970) and her coinvestigators were, for example. When working with audio or videotapes, the observers need not be working at the same time.

In fixed method research, double-coding by observers in order to assess interobserver reliability must be done independently. That is, the two observers should not be influenced in their ratings by seeing how the other observer has classified a particular event. Given this requirement, mechanical records have a clear advantage in that it is easy to make sure observers are not working together when they are coding.

After each observer has prepared his or her observation record, their records are compared to see how much agreement there is between them. If there is a high level of agreement, the system can be regarded as reliable. If agreement is not good, it may be because the operational definitions provided for the observers were not clear enough, because the observational task was too demanding (the observers were being expected to notice too many different things at once), or because the observer training was not adequate. Each of these problems is of course correctable. Unless the problem is fixed, the resulting observational data cannot be regarded as trustworthy or repeatable.

Observer agreement can be calculated in several ways. The simplest involves counting the number of observation intervals during which the observers agree and dividing that number by the total number of observation intervals in the study. This method is generally regarded as acceptable when whatever is being observed occurs about half the time. However, when the phenomenon to be observed occurs either very often or very rarely, this method will make it look like observers are agreeing very closely when in fact they may not be agreeing at all.

Consider, for example, the following situation. Suppose two observers observe a child for 20 observation intervals, and suppose the first observer records a single instance of the research phenomenon in the third interval, while the
second observer records a single instance of the observed phenomenon during interval 17. Computing their agreement by counting the number of agreements by the number of agreements and disagreements would produce a very high agreement index: They produced the same record 18 times (agreeing that the observation target was not seen in intervals 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, and 20) and different records twice (intervals 3 and 17). Their coefficient of agreement, then, would be 18 divided by 20, or 90 percent, but in reality they would have had absolutely no agreement at all on when the phenomenon occurred. Coefficients of rater agreement must be calculated in such a way as to genuinely reflect the degree of agreement between the two observers.

With observational data, interrater reliability is generally regarded as the most important form of reliability to demonstrate. When it has been demonstrated, the researcher can legitimately argue that the research phenomena as operationally defined are being observed similarly by two (or more) independent raters. Other forms of reliability, such as demonstrating consistency across times or across settings, may also be examined. Consistency in observation must be demonstrated and not assumed.

It is also just as important to demonstrate that a new observational system is valid as it is to demonstrate that any new data collection procedure is reliable. Reliability addresses the degree to which whatever is measured is repeatedly observable. Validity addresses whether whatever is measured has been defensibly understood. In other words, the validity of a measure relates to how data can be interpreted or to what sense can be made of the data.

When the conceptual and operational definitions guiding the observation have been carefully developed and articulated, observational data are generally regarded as easily having face and content validity (see chapter 11). Of course neither face nor content validity involves an empirical demonstration of the meaning of a score. However, criterion-related and more general construct validity must be empirically demonstrated for new observational systems just as they must for any new measurement procedure (see chapter 11).

Particularly if an observational system is to be used in additional studies by other researchers, it is important to demonstrate the criterion and/or construct validity of the method. Past demonstrations of interrater reliability, for example, only show that the system can be used reliably, not that it is being used reliably in the current work. Thus it is necessary to demonstrate rater reliability with each new scale application, in part because the observational system itself is only one component of the observational methodology. The observers are also an integral component of the methodology, and there is no guarantee that new observers will perform reliably using the same system other observers were able to use successfully.

It is also wise to collect evidence bearing on an observational system’s validity, and most well designed studies will afford these data as a matter of course. As will be recalled from the chapter on data collection in general, random error in measurement is inevitable. Therefore, it is often prudent to include multiple measures of key phenomena in any study. To the extent that important findings are replicated across different measures, a more compelling argument can be made for the apparently significant result.
Well-designed fixed method studies, then, often include more than one measure, or operationalization, for the central research phenomena. These several measures, while not properly interchangeable, should be meaningfully related; it follows that scores from them should be significantly correlated. To the extent that scores on an observational measure are correlated with scores on other measures of the same variable, the observational measure can be regarded as a valid indicator of the phenomenon under study.

**Participant Observation**

Participant observation is a flexible method of research widely used in sociology to examine social interactions, small groups, and organizational or community life. It involves the observer becoming a part of and a participant in the groups and communities being observed so that the role taken with those observed is less distant and more “inside.” Participant observation includes the same basic processes as seeing, recording, and coding or analyzing what is seen although adapted to the role the observer takes in the research setting. The Humphreys (1970) study and some parts of the Hochschild (1989) study, as well as the Rollins (1985) and Williams (1989, 1992) studies described in earlier chapters, are examples of participant observation research.

Observation, however, is often only one part of the participant observation process. For example, ethnomethodological studies, and others, typically include data gathering through both observation and intensive interviewing for the purpose of understanding the culture of a particular group or society. Even outside of formal interviewing, asking and listening take place as the observations are made. Documents and other artifacts may be collected at the site as well. Participant observation is above all opportunistic (Jorgensen 1989); therefore observation is not as clearly separated from other methods of data gathering as in other kinds of research.

Participant observation as a method is defined by the fact that the researcher is both a participant and actor in the situation being studied and an observer of it. However, the participant-observer role can be thought of as a continuum (Williamson et al. 1982). This continuum ranges from situations in which the observer is overtly identified as and acting almost always as a researcher with relatively little other interaction with those being observed (see Hochschild example above) to situations in which the observer is an active group participant, perhaps initially known as a group member and not as a researcher, whose research activities, such as record keeping, are designed to be covert or at least as invisible as possible in the setting.

Jorgensen (1989), for example, discusses three styles of participant observation: the active participant, who has a job or other social role in the setting in addition to the research; the privileged observer, who is known and trusted and has access to private information; and the limited observer, the most common one, who has no role other than researcher and who works to build trust in the setting over time. When multiple roles are occupied, as in the first two styles, managing and negotiating them can be a challenge.
In fact, major issues in any participant observation study concern how the observer will gain access to and enter the group or social system to be observed, whether and how the research is represented to others, and what role(s) the observer will engage in in the setting. In general, observations and other data gathering activities in the setting go from being very unfocussed in the beginning to being more focussed as data accumulate and potential answers to the research questions begin to emerge (Jorgensen 1989).

Participant observation research uses reactivity as a tool in the research. How those who are being observed react to the research and/or to the observer form part of what is learned about them, their social systems, and their view of the world. How the researcher reacts personally to those she observes are also data that inform her about the people and situations she studies. For example, like any traveler, a newcomer to a group or social setting will notice patterns of behavior that seem odd but that to a group member are so familiar and taken-for-granted they would never be commented on as unique or of any importance.

Like other forms of observation, participant observation is an expensive data collection method. It requires long periods of contact with the groups being studied, in part because over time initial atypical responses due to the presence of an outsider fade as the “outsider” becomes an “insider.” Although there is much to gain in knowledge from such familiarity, there is a risk in such long exposure that the researcher will begin adopting uncritically the views and perspectives of those who are studied and of failing to take adequate notice of things to which the observer is no longer naive.

During the course of a participant observation study, an observer is typically exposed to many sources of information and to many informants. Often, for example, a key informant will serve as a guide or a person who can help the researcher gain access to a particular group or social setting. Williams (1992) reports having such a “guide” in his study of crack houses in New York. This relationship requires careful nurture and can be quite rewarding to both parties. However, no one informant can be assumed to be typical of a group in every way, and no one informant can be assumed to be trustworthy and nonpartisan at all times and on all issues.

Triangulating data obtained on different occasions and from different sources is an important technique that can both validate information obtained and, when data are discrepant, suggest new dimensions of inquiry to pursue in order to explain apparent inconsistencies (Fetterman 1989). Seeking out and using multiple contacts and sources of information in the setting is thus an important step in determining which data are credible and worthy of report. Given the complexities involved in any individual entering, negotiating roles in, and leaving the field of study, the use of multiple data sources to validate information is much more common in participant observation than the use of multiple observers.

Recording the observations made is essential. Sometimes recording in the form of note-taking can go on in the setting, and sometimes it cannot. In either case, the researcher must plan adequate time for preparing the narrative records, generally called field notes, of what has been observed (Taylor & Bogdan 1984). These field notes should contain as much detail about who and where the recorded
events took place as well as about the events themselves as possible. There is always a great deal of detail observed in any interaction, and it may take a great deal of time before it is possible to distinguish important and relevant information from the unimportant. In fact, the analysis stage may be well underway by then. Recording data is generally less compelling and interesting than collecting it, but keeping the record complete and up-to-date is an essential discipline in participant observation research. While the notes must be as complete as possible in describing the events observed, they usually include interpretive material as well. Recording the subjective reactions of the observer brings them into awareness and makes them part of the data to be used in understanding the events studied. Thus observer and observed not only interact with each other; they are both “inside the frame” of the research itself.

Summary

Observation is involved whenever a recorder, human or mechanical, observes events for the purpose of collecting information about and making a permanent record of them. This is a broad definition, for observation is a versatile data collection method. Observation has been used as the data collection method both in highly structured forms of research and in flexible method studies in the form of participant observation. In fact, its use illustrates extremes in assumptions and styles of research. It can appear as the most “objective” or as the most “subjective” and reactive form of data gathering. Because visual data are so compelling, however, observational research has been essential to knowledge-building in many important areas of inquiry.

In some senses, observation captures the spirit of true empiricism. The greatest limitation of observation is probably practical: Done correctly, it is an expensive and time-consuming assessment methodology to use. Standing alongside this limitation is great strength: Done correctly it produces highly credible data especially when information about nonverbal and social behavior is needed.
Interviewing

The interview is so common in research that it is often taken for granted as a data collection method, but interviewing is in fact a complex, varied, and frequently misunderstood data collection technique. Despite these complexities, however, the research interview has certainly proven itself in use. In fact, it has been estimated that 90 percent of social science data have been obtained via interviews of one kind or another (Briggs 1986). However, as an eminent British sociologist observed, “Interviewing is rather like a marriage: everybody knows what it is, an awful lot of people do it, and yet behind each closed front door there is a world of secrets” (Oakley 1981:41). Recently insights from ethnographic interviewing and such disciplines as sociolinguistics have significantly enriched our understanding of the research interview as a data gathering method.

In the interview, a researcher and an informant meet in person or interact on the telephone, engaging in a conversation that has as its purpose the generation of data useful for research. This conversation may be a highly structured or an unstructured one; that is, it may be overtly guided almost wholly by the interviewer or by the informant. It may cover a wide range of information or only a narrow or specific topic. It may elicit in-depth content or relatively superficial information. The interview data may be recorded in summary on paper or verbatim on audiotape. Its purpose may be to collect information in order to examine preexisting hypotheses, to provide precise or in-depth description of selected phenomena, or to explore new or poorly understood phenomena.

Given the variety of its uses and of the forms that the research interview may take, the techniques used in interviewing also vary considerably. In fact, while there are some commonalities in all interviewing, it is really necessary to consider interviewing as it is done in fixed method and flexible method research separately. However, there are some commonalities that underlie all uses of the research interview. Knowledge of the strengths and weaknesses of interview techniques in...
general and of the various types of interviewing in particular will help in selecting the best data gathering method for a study and in evaluating studies that use interview methods.

**Uses and Advantages of the Interview**

Actions can be observed directly in the present; subjective experiences—motivations, feelings, meanings, and interpretations and memories of events from the past—cannot. However, these essential aspects of human behavior can be apprehended in words through interviewing. Written accounts may be used to capture these phenomena as well, but writing typically takes an informant more time and energy than speaking, and of course it requires literacy skills. People are often more willing and able to reflect at length on complex feelings, understandings, and past experiences through the spoken word than the written one.

In many ways, flexibility in the data gathering situation is the greatest strength of the interview. The degree of flexibility used will depend on whether the general approach to data gathering will be highly structured or unstructured. However, some degree of adaptability in the interaction is both necessary and desirable whatever form an interview takes. For the interviewer, there is always an element of being there with the respondent, which can afford the opportunity for observation above and beyond what the interview itself is designed to call forth.

The interview is an encounter and interaction between interviewer and interviewee. From the standpoint of the researcher, there are many advantages to being present as the informant encounters the interview questions. Pauses, hesitancies, and facial expressions may affect how a verbal answer should be interpreted. Outright questions from the interviewee or responses that seem “off the mark” may indicate that the interview questions have not been understood as intended. Unlike with the mailed questionnaire, often it is possible to interpret, explain, repeat, or redirect an interview question on the spot to obtain the information that the question was designed to elicit. Similarly, informants can be asked to elaborate on responses that are surprising or that seem incomplete, irrelevant, or hard to understand. Such flexibility in the moment may be greater or lesser depending on how structured the interviewing process is, but even in the most structured of interview approaches the interviewer can draw on observational data from the interaction to better understand the responses offered.

From the standpoint of the respondent, the presence of the interviewer may make it easier or harder to convey the information intended. With a skilled interviewer, the respondent will feel a sense of rapport, of being listened to and understood without judgement, and of focus on their own rather than the interviewer’s ideas, all of which can make it easier to talk. The interviewee also has a chance to “size up” the interviewer and the research itself through asking his or her own questions and discussing his or her own expectations about the interview and subsequent use of the data. Thus the research interview shares with many other kinds of professional encounters the quality of being a social interaction with a
defined purpose and its own social rules that are both like and unlike those of ordinary conversations.

Even in survey research using highly structured interview protocols, there are some situations in which interviewing conveys specific technical advantages. Data gathering in which the sequence or nature of the specific questions to be asked depends on the answers to previous questions, termed *contingency questioning,* is best handled through interviewing. Only a well-trained and experienced interviewer can reliably follow a complex sequence of questions through multiple contingencies. Respondents to questionnaires that require taking complex and different paths through the questions often have difficulty completing them, however clear the arrows and other written directions may be. Those who are interviewed, however, may never realize that questions that don’t apply to them have not been asked and that the interviewer is working hard to follow complex and differing paths through all the possible questions.

Similarly when exhaustive enumeration and coverage of a geographic sample or population is needed, there is no substitute for trained interviewers on the street for unit-to-unit canvassing of a given residential area, including identifying living spaces and respondents not previously known or officially recorded. The Gomez (1990) study of Cuban Americans in West New York, New Jersey, used survey interviewers to canvass the community in just this way. Only by going door to door in a specified block could all the households in which Cuban American residents lived be identified.

### Disadvantages of Interviews

As with all data gathering techniques, the advantages and disadvantages of the interview are closely related. Anonymity for the informant is not possible in a face-to-face interview. Confidentiality is usually offered, but once a person has been located and seen, his or her identity is known. This lack of anonymity is usually not a problem, but if issues of stigmatized or illegal activity should come up, it must be remembered that research interview data are not legally privileged and are certainly not immune from court subpoena. Any project in which data bearing on criminal matters or on mandated reporting issues such as child abuse might emerge must consider and deal with these limits on confidentiality (Herek et al. 1991, and see chapter 9 on ethics). Data generously given must be scrupulously recorded and safely used.

For the informant, an interview is a fairly intrusive form of data collection. It takes time, it may take place in the life space of the informant, and it may touch upon issues that are somewhat painful or difficult to discuss with a stranger. Interview procedures must deal thoughtfully with the impositions made on participants and must afford the informant the autonomy to decide whether or how completely to participate.

From the point of view of the interviewer, an interview is also a time-consuming endeavor. For each hour of contact with an informant, several hours of training, scheduling, and traveling before and of recording, data checking, and coding afterward may be necessary. Inevitably some interview appointments are cancelled and
must be rescheduled. This intensity of effort makes interviewing a relatively expensive form of data collection, especially if large samples are involved. Finally, while the interview encounter may result in a rapport that enhances the quality of the information obtained, factors known and unknown in the encounter between interviewer and respondent may affect the data in unpredictable ways. When face-to-face, interview responses can be affected by the race, gender, class, and appearance of the interviewer. In general, people may be most open in their responses when they perceive the interviewer to be someone like themselves. At other times, however, this may not be true; Herk et al. (1991) speculate, for example, that in AIDS research, gay male respondents who perceive their interviewers to be part of the gay and lesbian community may self-protectively underreport their unsafe sexual practices. Thus in survey research or other studies using multiple interviewers, care is often taken in the assignment of interviewers or in the management of their appearance and presentation to try to enhance the reliability of the data they gather. Even when there is only one interviewer, consideration must be given to all the factors in the interaction that might affect results.

The social rules of conversation can interfere with research interviewing. Respondents may only say what they think will be acceptable to the interviewer. For example, in conversation people prefer to express agreement rather than disagreement. “Yes, but . . .” answers may therefore be used in the research context to express a difference of opinion rather than a “no” response. The interviewer must then decide whether or not a “yes” or a “no” response was really intended from the point of view of the study.

**Interviewing in Flexible Method Research**

The interview designed to gather unstructured data, which includes what is termed clinical, focussed, qualitative, intensive, or ethnographic interviewing, is a common data gathering method used in flexible method research. In flexible method research, intensive interviewing is generally used because the research question springs from a desire to explore and learn more about some phenomenon that has not been previously studied or has been poorly understood. In this situation, the assumption is that the informant’s knowledge and experience of the phenomena of interest should guide the dialogue. Often the focus of flexible method study such as the psychological meaning of an event to a respondent is such that, by definition, it could only be captured in the respondent’s own words. Weiss (1994), for example, suggests using open-ended interviewing for developing detailed descriptions; eliciting multiple perspectives on an event, organization, or situation; understanding internal, interpersonal, or social processes; learning about how events are interpreted; and helping readers to understand a situation from a different point of view.

In ethnographic research, interviewing is often used in conjunction with participant observational techniques in order to elicit participants’ views and interpretations of the events that have taken place. Sometimes these narrative data are later reduced through content analysis either to a categorical or quantified
In flexible method research, the content of the interview itself is always developed as an integral part of the research process. Available instruments are not generally used because this would defeat the basic purpose of the research: to learn inductively about a phenomenon that was previously unknown or poorly understood. Often, however, an interview guide may be developed as a general map of the content and form of the questions to be asked (Padgett 1998; Weiss 1994).

Interviews in flexible method research vary in the extent to which control of the interview process and content is shared between interviewer and interviewee. At one extreme are ethnographic-style approaches in which the interviewee is treated as the expert who guides the interviewer toward increased knowledge of the cultural, social, or psychological phenomenon under study. In this way, the risk of imposing the researcher’s preconceptions on the phenomena reported is minimized. Such interviews will be as varied as the respondents who take part in them. The danger in such an approach is that the researcher will become so caught up in the view of reality offered up by the informant that he or she may adopt uncritically whatever typical or atypical rendition of reality the interviewee sets forth. However, even in ethnographic research, there are times when the researcher may choose to ask some preplanned or pointed questions, for example, to be sure to cover certain material, to check out inconsistencies in reports from different informants, or to explore emerging interpretations of the data. Patton (1980), for example, describes such approaches as “the informal conversational interview” as compared to the “general interview guide” (p. 197).

At the other extreme are interviews that are preplanned in almost all respects and conducted in the same way for each respondent even though the answers to the questions asked are unstructured ones. Patton (1990) terms this “the standardized open-ended interview” approach (p. 197). Care taken in advance about how each question is asked can yield useful data and may be an especially productive interviewing strategy to use when more than one interviewer will be used. The data gained from such similar interviews may also be more amenable to comparison between individuals and between groups of respondents. However, the danger in such an approach is that the researcher, in controlling the flow of the data, may miss some worthwhile information although it may reduce individual interviewer bias or variability in how questions are asked.

In fact, the interview may best be viewed as a special case of the conversation, an event in which two participants are to one extent or another mutually influencing the interaction and thus the data it will yield. In some interview situations, the interviewer will play the role of expert most of the time; in other instances, the respondent will be asked to play the role of expert predominantly, as in the life history interview. However, in no case must the interviewer fail to consider that her or his actions will wittingly or unwittingly shape the data the interview will yield. Verbal and nonverbal cues will either encourage an interviewee in a given line of conversation or discourage her. This influence cannot be willed or wished away. Rather it must be understood and used in a way that will enhance the credibility of the results.