From: Bernard, H. Russell. 2011. Research Methods in Anthropology: Qualitative and Quantitative Approaches. 5th edition. Lanham, MD. Rowman and Littlefield.Chapter 14.

# 14

# Direct and Indirect Observation

*You can observe a lot by just watching.*

—Yogi Berra 1964, cited in Berra and Garagiola 1998

Interviewing is a great way to learn about attitudes and values. And it's a great way to find out what people think they do. When you want to know what people *actually do*, however, there is no substitute for watching them or studying the physical traces their behavior leaves behind. This chapter is about **direct observation** (watching people and recording their behavior on the spot) and **indirect observation** (the archeology of human behavior).

There are two big strategies for direct observation of behavior. You can be blatant about it and **reactive**, or you can be unobtrusive and **nonreactive**. In reactive observation, people know that you are watching them and may play to their audience—you. You can wind up with data about what people want you to see and learn little about what people do when you're not around. In unobtrusive observation, you study people's behavior *without their knowing it*. This stops people from playing to an audience, but it raises tough ethical questions. We'll get to some of those problems later in this chapter.

We begin with the two most important methods for direct observation, **continuous monitoring** and **spot sampling** of behavior. Then we take up unobtrusive observation (and the ethical issues associated it), and, finally, indirect observation.

## CM—CONTINUOUS MONITORING

In continuous monitoring, or CM, or **focal follows**, you watch a person, or group of people, and record their behavior as faithfully as possible. The technique was developed in the field of management by Charles Babbage, the 19th-century mathematician who invented the computer. He studied the behavior of workers in a factory and determined that a pound of number 11 straight pins (5,546 of them) should take exactly 7.6892 hours to make (Niebel 1982:4; original: Babbage 1835:184).

CM is widely used in assessing the quality of human interactions—between, for example, adolescent girls and their mothers (Baril et al. 2009), workers and employers (Sproull 1981), the police and civilians (Sykes and Brent 1983), clinical professors and young physicians (Graffam et al. 2008) (**Further Reading:** continuous monitoring).

CM is the core method of **ethology** (Hutt and Hutt 1970; Lorenz 1981). Most ethologists study nonhuman animals (everything from moths to fish to chimpanzees), but Darwin (1998 [1872]) used direct observation of facial expressions to study emotions in humans and animals—an area of interest ever since (Ekman 1973, 1980; Leeland 2008). CM is a mainstay in behavioral psychology for assessing anxieties and phobias (Harb et al. 2003), and it has been used to study how people eat (Stunkard and Kaplan 1977; Zive et al. 1998) and how people use architectural space (Bechtel 1977). CM is a staple method in the study of how hunters and fishermen make a living (Aswani 2005; Bird et al. 2009; Hawkes et al. 1991; Koster 2007) and how children learn to hunt and forage (Hewlett and Lamb 2005). CM is one of the all-around varsity methods **(Further Reading:** ethology and human ethology).

## ETHOGRAMS

It is standard practice in ethology to develop an **ethogram**, or list of behaviors, for a species being studied. It's painstaking work. Lee and Brewis (2009) spent a summer doing pilot research to develop a list of 37 behaviors associated with foraging by children in a Mexican shantytown. Then they followed 20 children for a total of 15 hours each (watching each child for three blocks of about 5 hours at a time) coding for everything in their list of behaviors. The codes included things like begging, getting a gift of food from a peer, and getting money through informal employment. Some of the behaviors were: being in school, doing chores inside the house, and being en route to or from school or work and household. Figure 14.1 shows a part of one of their observations.

| | |
|---|---|
| 12:00 | Father gives Carlos 1 peso |
| 12:15 | Leaves to play kites with friends, on way to the bakery where he works informally |
| 12:34 | Given a bakery basket to carry to the next neighborhood to sell door-to-door |
| 13:09 | Forages a medium orange from a neighborhood tree, and eats it while carrying the bread door-to-door (does not eat any of the bread) |
| 13:20 | Returns basket to baker, and is paid 4.5 pesos |
| 13:29 | Returns home, watches television |
| 14:04 | Gives his father 50 centavos |
| 14:05 | Goes to corner store alone, and uses remaining 5 pesos to purchase flavored corn chips (120 grams) and Tic-Tac candy (16 gram pack) |
| 14:09 | Shares chips and candy with his nephew (age 5 years) and friend (11 years), until all is eaten |
| 14:30 | Returns home |

FIGURE 14.1.
Sequence of activities of ''Carlos,'' age 11, Thursday, July 15, 2004 (a sunny day).
SOURCE: Lee and Brewis (2009:441).

Kneidinger et al. (2001) studied touching behavior in 119 mostly white, male baseball players and 52 mostly white, female softball players in six major universities in the southeastern United States. Kneidigner et al. developed an ethogram of 37 touching behaviors before they even launched their main study. The touching behaviors included things like tapping gloves, high fives, butt slaps, and chest grabs (''one participant grabs the front of the other participant's shirt'').

The main study involved watching and recording 1,961 touching behaviors across 99 innings of baseball for the men and 1,593 touching behaviors across 63 innings of softball for the women. Among the interesting results: Men and women touched each other the same amount after winning games, but women touched each other more than men touched each other after losing games (Kneidinger et al. 2001:52).

## CONTINUOUS MONITORING IN ANTHROPOLOGY

CM has a long and noble history in anthropology. When Eliot Chapple was still a student at Harvard in the 1930s, he built a device he called the ''interaction chronograph'' for recording on a rolling sheet of paper the minute features (facial expressions, gestures) of human interaction. The interaction chronograph is, as far as I can tell, the unheralded forerunner of the hand-held computer recording systems used for continuous monitoring in ethology, psychology, and anthropology today (Chapple 1940; Chapple and Donald 1947).

In 1949, John Roberts and a Zuni interpreter took turns sitting in one of the rooms of a Zuni house, simply dictating their observations into a tape recorder. (That recorder, by the way, was the size of a suitcase and weighed 30 pounds.) This went on for 5 days and produced data for a 75,000-word book, rich in detail about everyday Zuni life. Figure 14.2 shows some excerpts from Roberts's work.

People let Roberts park in their homes for 5 days because Roberts was a participant

0940

E1DaE1S09 is dressed in blue denim overalls and blue denim shirt. FaSiSoSo is wearing a cotton shirt, heavy trousers, jacket and oxfords. The girls are wearing dresses, socks, and shoes. 2Da24 has on a blouse, skirt, green socks, and oxfords.

0941-(FaSiSo37D)

YoDaSo1 came into SCR from ESCR carrying a little toy in his hand.

0945-(FaSiSo37d)

I intended going to the buck herd today to take out my bucks (rams). I was going to bring them down to Zuni to feed them to get them in good shape—but there is no time to go over there today. I think I will go tomorrow.

AdE1So27A went into ESCR, ENCR, and SER, but he had nothing to report.

Mo61 is still in SER shelling corn.

0950

Mo61 walks back into WNCR to build a fire in the WNCR cooking stove.

AdE1So27A says that she is going to make hominy with the stew.

3Da22 is mounting turquoise on sticks for grinding.

YoDaSo1 came into SCR a few moments ago with a homemade cardboard horse which had been cut out by YoDaHu22.

2Da2Da3 followed YoDaSo1.

This house is full of activity and the children are running back and forth. They are not playing outside today because the weather is poor.

E1Da28 is mounting turquoise on sticks in preparation for grinding. She has a fire going in WR, which is a very large room to heat.

FIGURE 14.2.

Excerpts from Roberts's observations of a Zuni household. Persons and things are identified by shorthand notation. For example, 2Da2Da3 is the family's second daughter who is 3 years old. Sequence begins at 9:40 A.M. and ends at 10:00 A.M.

SOURCE: J. M. Roberts, *Zuni Daily Life.* © 1965 [orig. 1956], HRAF Press. Reproduced with permission.

**Table 14.1 Nutritional Behavior of Four Pilagá Indian Children**

| Behavior | *Yorodaikolik (4 years)* | *Tanpani (8–9 years)* | *Naicho (6 years)* | *Deniki (15 months)* |
|---|---|---|---|---|
| Given food | 9 | 7 | 27 | 42 |
| Deprived of food | 7 | 0 | 2 | 4 |
| Deprivation followed by restitution | 1 | 0 | 1 | 2 |
| Attempt made to deprive of food | 2 | 1 | 1 | 4 |
| Gives food | 8 | 7 | 1 | 3 |
| Withholds food | 6 | 5 | 2 | 3 |
| Deprives others of food | 2 | 6 | 1 | 0 |
| Attempts to deprive others of food | 1 | 5 | 0 | 1 |
| Receives part of a whole | 0 | 1 | 0 | 4 |
| Punished while eating | 0 | 0 | 0 | 3 |
| Total observations of each child from which these are taken | 190 | 207 | 238 | 208 |

SOURCE: I. N. Mensh and J. Henry. 1953. Direct Observation and Psychological Tests in Anthropological Field Work. *American Anthropologist* 55(4): 466. http://www.anthrosource.net.

observer of Zuni life and had gained his informants' confidence. Even earlier, in 1936–37, Jules and Zunia Henry did fieldwork among the Pilagá Indians of Argentina. Among the data they collected was a set of direct observations of children. Table 14.1 shows the data from observations made on four children for 10 kinds of behaviors associated with eating and food sharing.

The data in table 14.1 were extracted from 843 observations of children's behavior. Here are two of those observations from the original data:

> The three children of Diwa'i are feeding peacefully together. Deniki, the baby, waves his hand for food and mother gives him a small piece of palm dipped in fat. After eating a second piece he is given the breast.
>
> Deniki, Nacho, and Soroi are together. Deniki is holding a dish with a very small quantity of cooked fruit in it. Soroi says, "Share it with me," and takes one fruit out of the dish. Naicho immediately snatches another one away violently, but not before Deniki has already taken one out, which he then offers to Naicho, appearing not to comprehend her action. (Mensh and Henry 1953:467)

### The Zapotec Children Study

Douglas Fry used CM to study aggressive play among Zapotec children. From 1981 to 1983, Fry did 18 months of participant observation fieldwork in La Paz and San Andrés, two small Zapotec-speaking villages just 4 miles apart in the Valley of Oaxaca, Mexico. During the last 5 months of his research, Fry did direct, continuous monitoring of 24 children (3–8 years old) in each village. Before that, he visited almost all the households in the villages several times so that children had become accustomed to him when he began his intensive observation.

Fry describes his data collection procedures clearly:

> The formal focal sampling observations were conducted between May and September of 1983. They represent each day of the week and encompass the daylight hour. Most observations (84%) were conducted within family compounds, although children were also observed in the streets, town squares, school yards, fields, and hills. I alternated sampling between the two communities on a weekly to biweekly basis. A total of 588 observations were conducted, resulting in an average of approximately 12 observations

for each focal child (M = 12.25, *SD* = 6.21). On average, each focal child was observed for just over 3 hours (M = 3.13 hours, *SD* = 1.39 hours), resulting in a total of 150 hours of observation time for the entire sample. [It is common in scientific papers to report means and standard deviations; hence the M and *SD* figures in this paragraph. HRB]

Focal observations were narrated into a tape recorder carried in a small backpack or recorded on paper using a shorthand system. I recorded a running commentary of the behaviors engaged in by the focal child, using behavior elements defined in the previously developed ethogram. I moved with a focal child in order to maintain continuous visual contact (Altmann 1974), but did not remain so close as to interfere with actions or unduly attract the child's attention. Whenever a focal child engaged in any type of antagonistic behavior, the specifics of the interaction were noted, including identity of the interactant(s) and any facial expressions or gestures. For instance, interactions such as the following were recorded: Focal boy punches, pushes sister of 3 year old while laughing (sister does nothing in response). (Fry 1990:326–27)

Fry developed his ethogram of Zapotec children by watching them in public places before beginning his study of focal individuals. Based on 150 hours of focal child observation, Fry's data contain 764 episodes of what he calls "play aggression" and 85 episodes of "serious aggression."

Play aggression is a punch, kick, tackle, etc., accompanied by smiles, laughs, and playfaces. Serious aggression acts are episodes accompanied by low frowns, bared teeth, fixated gazes, and crying. Fry found that when girls initiated serious aggression, it was almost always with other girls (93% of cases). But when boys initiated serious aggression, it was just as likely to be with girls as with other boys (**Further Reading:** studying children under natural conditions) (box 14.1).

**BOX 14.1**

## DIRECT OBSERVATION VERSUS SELF-REPORTS

Pearson (1990) used CM to study how Samoans in Western Samoa, American Samoa, and Honolulu used energy. Pearson had the idea that, as Samoans moved to Honolulu and became more urbanized, there would be changes in lifestyle and that these changes would show up in their energy intake and expenditure. Pearson asked people to recall their activities over the past 24 hours, and then, to check the **24-hour recall** data, he and a female assistant monitored 47 men and 43 women.

The team did 825 hours of observation and had their subjects in direct view 92% of the time. The estimates of energy expenditure from direct observation of the 47 men were 33%–80% lower than the estimates from the recall data. For women, the estimates were 27%–53% lower. Women had better recall of their activities, but men and women were both way off the mark, particularly in recalling their light-to-moderate work of the previous day. Pearson's work makes it clear that, when it comes to measuring energy expenditure, recall is not a good substitute for observation.

### The Shopping Study

Martin Murtaugh (1985) used CM to study the use of arithmetic by grocery shoppers. He recruited 24 adults in Orange County, California. Accompanied by two observers, each informant wore a tape recorder while shopping at a supermarket. As the informants went about their shopping, they talked into the tape recorder about how they were deciding which product to buy, what size to choose, and so on.

One observer mapped the shopper's route through the store and recorded the prices and amounts of everything purchased. The other researcher kept up a running interview with the shopper, probing for details. Murtagh was aware of the potential for reactivity in his study. But he was interested in understanding the way people thought through ordinary, everyday arithmetic problems, and his experiment was a good way to generate those problems under natural conditions.

Many CM researchers record their own observations orally. It's less tedious than writing; it lets you focus your eyes on what's going on; it lets you record details later that might be left out of a on-the-spot written description; it avoids the limitations of a check list; and it lets you get information about context as well as about the behavior you're studying. Moreover, you can easily transcribe your recorded observations, once you've got your voice recognition software trained (see above, chapter 8, and appendix E).

But there are trade-offs. If you want measurements from qualitative data (like running commentaries on tape), you have to code them. That is, you have to listen to the recordings, over and over again, and decide what behaviors to code for each of the people you observe. Coding on the spot (by using a behavioral checklist or by inputting codes into a handheld computer) produces immediate quantitative data. You can't code *and* talk into a recorder at the same time, so you need to decide what kind of data you need and why you need them before you choose a method.

If you are trying to understand a behavioral *process*, then focus on qualitative data. If you need measurements of *how much* or *how often* people engage in this or that behavior, then focus on quantitative data. And, as always, who says you can't do both?

## CODING CONTINUOUS MONITORING DATA

Go to a shopping mall and record the interaction behavior of 30 mother-child pairs for 2 minutes each. Record carefully the number of children each mother has and her interaction with each child. Try to find out whether interaction patterns are predictable from: (1) the number of children a mother has to cope with; (2) the ages of the children; (3) the socioeconomic class or ethnicity of the family; or (4) some other factors.

This exercise is instructive, if not humbling. It's a real challenge to code for socioeconomic class and ethnicity when you can't talk to the people you observe. Do this with at least one colleague so you can both check the reliability of your coding.

In hypothesis-testing research, where you already know a lot about the people you are studying, you go out to observe armed with a coding scheme worked out in advance. The idea is to record any instances of behavior that conform to the items in the scheme. This allows you to see if your hunches are correct about conditions under which certain behaviors occur. In some studies, you might be interested in noting instances of aggressive versus submissive behavior. In other cases, those variables might be irrelevant.

### Coding Schemes

Just as with attitude scales and surveys (in chapters 9 and 11), there's no point in reinventing the wheel. Over the years, researchers have developed coding schemes for using direct observation in many different situations—in studies of interactions between married couples, in studies of teacher effectiveness, in worker-management negotiations, and so on. If others have developed and tested a good system for coding behaviors of interest to you, use it. Don't feel that it's somehow more prestigious or morally better for you to make up everything from scratch. Knowledge grows when researchers can compare their data to the data others have collected using the same or similar instruments.

Figure 14.3 shows the basic coding scheme for **interaction process analysis**, a system developed 60 years ago by Robert F. Bales in his research on communications in small groups (Bales 1950).

| PROBLEM AREAS | | | OBSERVATION CATEGORIES |
|---|---|---|---|
| Positive Reactions | A | 1 | Shows solidarity, raises other's status, gives help, rewards |
| | | 2 | Shows tension release, jokes, laughs, shows satisfaction |
| | | 3 | Agrees, shows passive acceptance, understands, concurs, complies |
| Attempted Answers | B | 4 | Gives suggestions, direction, implying autonomy for other |
| | | 5 | Gives opinion, evaluation, analysis, expresses feelings, wishes |
| | | 6 | Gives orientation, information, repeats, clarifies, confirms |
| Questions | C | 7 | Asks for orientation, information, repetition, confirmation |
| | | 8 | Asks for opinion, evaluation, analysis, expression of feeling |
| | | 9 | Asks for suggestions, direction, possible ways of action |
| Negative Reactions | D | 10 | Disagrees, shows passive rejection, formality, withholds help |
| | | 11 | Shows tension, asks for help, withdraws out of field |
| | | 12 | Shows antagonism, deflates other's status, defends or asserts self |

FIGURE 14.3.
Categories for direct observation.
SOURCE: SOURCE: R. F. Bales, ''Some Uniformities of Behavior in Small Social Systems.'' In *Social Interaction Systems: Theory and Measurement*, p. 165. New Brunswick, NJ: Transaction Publishers, 2000.

Despite its age, the Bales coding scheme continues to be used in the study of classrooms (Koivusaari 2002) and work teams (Nam et al. 2009)—in fact, in any situation where people interact with one another.

Stewart (1984) audiotaped 140 doctor-patient interactions in the offices of 24 family physicians and assessed the interactions with Bales's interaction process analysis. Ten days later, Stewart interviewed the patients at their homes to assess satisfaction and compliance. That is, were the patients satisfied with the care they'd gotten and were they taking the pills they'd been told to take? Sure enough, when physicians are coded as engaging in many patient-centered behaviors, patients report higher compliance and satisfaction.

One of the best things about the interaction process analysis system is that any act of communication can be identified as being one of those 12 categories in figure 14.3, and the 12 categories are recognizable in many cultures around the world. A more detailed outline for coding interpersonal relations was developed by Bales and Cohen (1979). A complete course on how to use their system is available in their book, aptly titled *SYMLOG*, which stands for ''systematic multiple level observation of groups'' (box 14.2).

BOX 14.2

### DOING CONTINUOUS MONITORING IS HARDER THAN IT SOUNDS

Make no mistake about this: Continuous monitoring is tough to do. It takes several months of intensive training for observers to become adept at using complex coding schemes. In the 1920s, Rudolf von Laban developed a system of 114 signs with which to record dance or any other set of human movements (see Lange 1975). If you are trying to understand the meaning of complex human body movement, you've got to start with some way to record the basic data, like any other text. Brenda Farnell used Laban's script (known as **labanotation)** in her meticulous research on Plains Indian sign language (1995) and has written extensively about how the system can be used in the anthropological study of human movement in general (Farnell 1994, 1996; Farnell and Graham 1998).

These days, behavioral coding in psychology and ethology is being done with hand-held computers and software that lets you program any key to mean ''initiates conversation,'' ''reciprocates affect,'' or whatever. (See Ice [2004] and Koster [2006] for details about using this technology and see appendix E.) As it becomes easier for fieldworkers to observe and code behavior at the same time, I think we'll see renewed interest in continuous monitoring and in the use of complex coding schemes like labanotation.

One of the problems in the use of direct observation is the need for reliable coding by several researchers of the same data. We'll take up measures of intercoder reliability in chapter 19 on text analysis. The problem of testing intercoder reliability is the same, whether you're coding text or behavior.

## COMPARATIVE RESEARCH—THE SIX CULTURE STUDY

Broad, general coding schemes are particularly useful for comparative research. Whether you're comparing sessions of psychotherapy groups, interaction sessions in laboratory experiments, or the natural behavior of people in field studies, using a common coding scheme really pays off because you can make direct comparisons across cases and look for generalizations.

The most important comparative study of children ever was run by Beatrice and John Whiting between 1954 and 1956. In the Six Culture Project, field researchers spent from 6 to 14 months in Okinawa, Kenya, Mexico, the Philippines, New England, and India. They made a total of some 3,000 5-minute (continuous monitoring) observations on 67 girls and 67 boys between the ages of 3 and 11.

Observations were limited to just 5 minutes because they were so intense, produced so much data, and required so much concentration and effort that researchers would have become fatigued and lost a lot of data in longer sessions. The investigators wrote out, in clear sentences, everything they saw children doing during the observation periods. They also recorded data about the physical environment and others with whom children were interacting.

The data were sent from the field to Harvard University for coding according to a scheme of 12 behavior categories that had been worked out in research going back some 15 years before the Six Culture Study began. The behavioral categories included: seeks help, seeks attention, seeks dominance, suggests, offers support, offers help, acts socially, touches, reprimands, assaults sociably, assaults not sociably, symbolic aggression (frightens, insults, threatens with gesture, challenges to compete). (Full details on the use of the Whiting scheme are published in Whiting et al. [1966]. See Whiting and Whiting [1973] for a discussion of their methods for observing and recording behavior.)

On average, every 10th observation was coded by two people, and these pairs of "coding partners" were rotated so that coders could not slip into a comfortable pattern with one another. Coders achieved 87% agreement on children's actions; that is, given a list of 12 kinds of things a child might be doing, coders agreed 87% of the time. They also agreed 75% of the time on the act that precipitated a child's actions and 80% of the time on the effects of a child's actions (Whiting and Whiting 1975:55).

The database from the Six Culture Study consists of approximately 20,000 recorded acts, for 134 children, or about 150 acts per child, on average.

Very strong conclusions can be drawn from this kind of robust database. For example, Whiting and Whiting (1975:179) note that nurturance, responsibility, success, authority, and casual intimacy "are types of behavior that are differentially preferred by different cultures." They conclude that "these values are apparently transmitted to the child before the age of six." They found no difference in amount of nurturant behavior among boys and girls 3–5 years of age. After that, however, nurturant behavior by girls increases rapidly with age, while boys' scores on this trait remain stable.

By contrast, reprimanding behavior starts out low for both boys and girls and increases with age equally for both sexes, across six cultures. The older the children get, the more likely they are to reprimand anyone who deviates from newly learned cultural rules. "Throughout the world," the Whitings conclude, "two of the dominant personality traits of children between seven and eleven are self-righteousness and bossiness" (1975:184). Anyone who grew up with an older sibling already knows that, but the Whitings' demonstration of this cross-cultural fact is a major scientific achievement.

## USING VIDEO FOR CONTINUOUS MONITORING

Even with a fixed coding scheme, an observer in a CM situation has to decide among alternatives when noting behavior—whether someone is acting aggressively, or just engaging in rough play, for example. Recording behavior on film or video lets several analysts study the behavior stream and decide at leisure how to code it. It also makes your data available for coding by others, now and in the future. (Human ethologists, like Irenäus

Eibl-Eiblsfeldt [1989], have amassed hundreds of miles of film and videotape of ordinary people doing ordinary things across the world.)

In the 1970s, Marvin Harris and his students installed videotape cameras in the public rooms of several households in New York City. Families gave their permission, of course, and were guaranteed legal control over the cameras during the study and of the videotapes after the cameras were removed. Teams of observers monitored the equipment from remote locations. Later, the continuous verbal and nonverbal data were coded to study regularities in interpersonal relations in families.

Anna Lou Dehavenon (1978), for example, studied two black and two white families for 3 weeks and coded their nonverbal behavior for such things as compliance with requests and the distribution and consumption of foods in the households. Dehavenon's data showed that the amount of authoritarianism in the four families correlated perfectly with income differences. The lower the family income, the more superordinate behavior in the home (1978:3).

One would hypothesize, from participant observation alone, that this was the case. But *testing* this kind of hypothesis requires the sort of quantified data that straightforward, direct observation provides. (See Sharff [1979] and Reiss [1985] for two more studies of households using the Harris videotapes.)

By the 1980s, anthropologists were using video in studies of consumer behavior. Observers at Planmetrics, a marketing research firm, videotaped 70 volunteer parents, for over 200 hours, as the volunteers diapered their babies. The research was done on contract with Kimberly-Clark, manufacturer of "Huggies," a brand of disposable diapers. The cameras were not hidden, and after a while people just went about their business as usual, according to Steven Barnett, the anthropologist who led the study.

Close observation showed that many parents could not tell whether their babies needed a diaper change, so the researchers recommended that the diapers contain an exterior chemical strip that changed color when the baby was wet. The observers also noticed that parents were powdering their babies' legs and that parents were treating the red marks left by the diaper gathers as if the marks were diaper rash. The firm recommended that the gathers be redesigned so that there would be no more red marks (Kilman 1985; Lewin 1986). Today, video is used routinely in research on product design and use (Wasson 2000) (box 14.3).

### CM and Reactivity

Finally, there are two ways to lower reactivity in continuous monitoring. One of them is participant observation. Once you've built up rapport and trust in a field situation, people are less likely to change their behavior when you're around. Even if they do change their behavior, you're more likely to notice the change and take that into account.

The second way to lower reactivity is training. We can't eliminate observer bias entirely, but lots and lots of evidence shows that training helps make people better—more reliable and more accurate—observers (Hartmann and Wood 1990; Kent et al. 1977). We do the best we can. Just because a "perfectly aseptic environment is impossible," Clifford Geertz (1973:30) reminds us (paraphrasing the economist Robert Solow 1970:101), doesn't mean we "might as well conduct surgery in a sewer."

Joel Gittelsohn and his coworkers (1997) tested the effects of participant observation and training on reactivity in their study of child-care practices in rural Nepal. Over the course of a year, 10 trained fieldworkers observed behavior in 160 households. Each home was visited seven times. Except for a 3–4 hour break in the middle of the day, the field-

**BOX 14.3**

## VIDEO IS EASIER THAN EVER

As video cameras have gotten smaller, easier to use, and less expensive, more field researchers have been using this technology for close examination of behavior streams. Brigitte Jordan, for example, used videotape in her study of birthing events across cultures (1992; see also Jordan and Henderson 1993) and Kremer-Sadlik and Paugh 2007) documented the emergence of what their middle-class informants in Los Angeles called ''quality time'' in everyday family life. You can code video today as easily as you can code written text. More about this in chapter 19. Look for lots more use of systematically recorded and coded video in anthropology.

workers observed a focal child, 2–5 years of age, and all the caregivers of that child, from 6:00 A.M. until 8:00 P.M. This study, then, involved both children and adults.

The observers coded for over 40 activities, including health-related behaviors, feeding activities, and various kinds of social interactions (punishment, affection, and so on). The rate of some behaviors changed a lot over the course of the year. On average, across 1,101 observations, the number of times per day that a caregiver served food to a child without asking the child if he or she wanted it fell by half.

The observers also coded *each time they were interrupted* by one of the people whom they were observing (and what the interruption was about: e.g., light conversation, being asked for favors or medicine). This allowed Gittelsohn et al. to track reactivity across the seven household visits. Reactivity was noticeable during the first visit and then fell off dramatically. This study shows clearly that: (1) reactivity exists, and (2) it goes away quickly when indigenous observers stay on the job over time (Gittelsohn et al. 1997).

## SPOT SAMPLING AND TIME ALLOCATION STUDIES

**Instantaneous spot sampling**, or **time sampling**, was developed in behavioral psychology in the 1920s and is widely used in ethology today. In **time allocation** (TA) studies, which are based on time sampling, an observer appears at randomly selected places, and at randomly selected times, and records what people are doing when they are first seen (Gross 1984).

The idea behind the TA method is simple and appealing: If you sample a sufficiently large number of representative acts, you can use the percentage of *times* people are seen doing things (working, playing, resting, eating) as a proxy for the percentage of *time* they spend in those activities.

Charles Erasmus used spot sampling in his study of a Mayo Indian community in northern Mexico (1955). As Erasmus and his wife went about the village, investigating "various topics of ethnographic interest," they took notes of what people were doing at the moment they encountered them. They did not use a representative sampling strategy but they were very systematic in their recording of data.

> Individual charts were made for each man, woman, and child in the village, and on those charts were noted the page numbers from the field log where the activity descrip-

> tions were to be found. These page numbers were recorded on the charts according to the hours of the day when the observations were made. Thus, the individual charts served as indexes to the field log as well as a means of making sure that equal attention was being given to all families at all hours of the day. Periodic examination of the charts showed which households and which hours of the day were being neglected, so that visits about the community could be planned to compensate for these discrepancies. (Erasmus 1955:325)

It's difficult to top this research for sheer elegance of design and the power of the data it produced. In the 3 months from July to September 1948, the Erasmuses made about 5,000 observations on 2,500 active adults, 2,000 children, and 500 elders in the community. From those observations, Erasmus demonstrated that men in the village he studied spent about the same time at work each day as did semiskilled workers in Washington, DC. At the time, Melville Herskovits was trying to combat the racist notion that primitive and peasant peoples are lazy and unwilling to exert themselves. Herskovits's assertion was vindicated by Erasmus's TA research.

### Reactivity in TA Research

In CM, getting around the reactivity problem involves staying with the program long enough to get people accustomed to your being around. Eventually, people just get plain tired of trying to manage your impression and they act naturally.

In TA research, the trick is to catch a glimpse of people in their natural activities before they see you coming on the scene—before they have a chance to modify their behavior.

Richard Scaglion (1986) did a TA survey of the residents of Upper Neligum, a Samakundi Abelam village in the Prince Alexander Mountains of East Sepik Province in Papua New Guinea. "It is not easy," he says, "for an anthropologist in the field to come upon an Abelam unawares. Since I did not want to record 'greeting anthropologist' as a frequent activity when people were first observed, I often had to reconstruct what they were doing immediately before I arrived" (p. 540).

Monique Borgerhoff-Mulder and Tim Caro (1985) coded the observer's judgment of whether people saw the observer first, or vice versa, and compared that to whether the Kipsigis (in Kenya) they were studying were observed to be active or idle. People were coded as being idle significantly more often when they spied the observer coming before the observer saw them.

Did people become idle when they saw an observer approaching? Or was it easier for idle people to see an observer before the observer saw them? Borgerhoff-Mulder and Caro found that people who were idle were sitting or lying down much more often than were people who were active. People at rest may be more attentive to their surroundings than those who are working and would be judged more often to have seen the researcher approaching.

## SAMPLING PROBLEMS

There are five questions to ask when drawing a sample for a TA study:

1. Who do I watch?
2. Where do I go to watch them?
3. When do I go there?
4. How often do I go there?
5. How long do I spend watching people when I get there? (Gross 1984)

Allen Johnson's study (1975) of the Machiguenga is instructive. The Machiguenga are horticulturalists in the Peruvian Amazon. They live along streams, in small groups of related families, with each group comprising from about 10 to 30 people, and subsist primarily from slash-and-burn gardens. They supplement their diet with fish, grubs, wild fruits, and occasional monkeys from the surrounding tropical forest. Johnson spent 14 months studying the Machiguenga in the community of Shimaa.

Johnson's strategy for selecting people to study was simple: Because all travel was on foot, he decided to sample all the households within 45 minutes of his own residence. This produced a convenience sample of 13 households totaling 105 persons. The Machiguenga live along streams, so each time Johnson went out he walked either upstream or downstream, stopping at a selected household along the route. He selected the hour of the day to go out and the houses to vist at random.

Thus, Johnson used a nonrandom sample of all Machiguenga households, but he randomized the times that he visited any household in his sample. This sampling strategy sacrificed some external validity, but it was high on internal validity. Johnson could not claim that his sample of households *statistically* represented all Machiguenga households. His 14 months of experience in the field, however, makes his claim for the representativeness of his data credible.

That is, if Johnson's data on time allocation in those 13 households seem to *him* to reflect time allocation in Machiguenga households generally, then they probably do. But we can't be sure. Fortunately, randomizing his visits to the 13 households, and making a lot of observations (3,945 of them, over 134 different days during the 14-month fieldwork period), gives Johnson's results a lot of *internal* validity. So, even if you're skeptical of the external validity of Johnson's study, you could repeat it (in Shimaa or in some other Machiguenga community) and see whether you got the same results.

Regina Smith Oboler (1985) did a TA study among the Nandi of Kenya. She was interested in differences in the activities of adult men and women. The Nandi, Oboler said, ''conceptualize the division of labor as sex segregated. Is this true in practice as well? Do men and women spend their time in substantially different or similar types of activities?'' (p. 203).

Oboler selected 11 households, comprising 117 people, for her TA study. Her sample was not random. ''Selecting a random sample,'' she said, ''even for one *kokwet* (neighborhood) would have made observations impossibly difficult in terms of travel time'' (Oboler 1985:204). Instead, Oboler chose a sample of households that were matched to social and demographic characteristics of the total population and within half an hour walking distance from the compound where she lived.

Oboler divided the daylight hours of the week into 175 equal time periods and gave each period (about 2 hours) a unique three-digit number. Then, she chose time periods at random from the list of 175 numbers to visit each household. She visited each household four times a week (on different days of the week) during 2 weeks each month and made nearly 1,500 observations on those households during her 9 months in the field.

Oboler found that, for her sample of observations, adult men spend around 38% of their time ''in activities that might reasonably be considered 'work' by most commonly used definitions of that term'' (Oboler 1985:205). Women in her sample spent over 60% of their time working.

#### Sampling Table for TA Studies

Table 14.2 shows the number of spot observations necessary to estimate the frequency of an activity to within a fractional accuracy. It also tells you how many observations you need if you want to see an activity at least once with 95% probability.

**Table 14.2 Number of Observations Needed to Estimate the Frequency of an Activity to within a Fractional Accuracy**

| *True frequency of activity* | *Number of observations needed to see the activity at a particular fraction of accuracy* | | | | | | | *To see activities at least once with 95% probability* |
|---|---|---|---|---|---|---|---|---|
| *f* | 0.05 | 0.10 | 0.15 | 0.20 | 0.30 | 0.40 | 0.50 | |
| 0.01 | 152127 | 38032 | 16903 | 9508 | 4226 | 2377 | 1521 | 299 |
| 0.02 | 75295 | 18824 | 8366 | 4706 | 2092 | 1176 | 753 | 149 |
| 0.03 | 49685 | 12421 | 5521 | 3105 | 1380 | 776 | 497 | 99 |
| 0.04 | 36879 | 9220 | 4098 | 2305 | 1024 | 576 | 369 | 74 |
| 0.05 | 29196 | 7299 | 3244 | 1825 | 811 | 456 | 292 | 59 |
| 0.06 | 24074 | 6019 | 2675 | 1505 | 669 | 376 | 241 | 49 |
| 0.07 | 20415 | 5104 | 2268 | 1276 | 567 | 319 | 204 | 42 |
| 0.08 | 17671 | 4418 | 1963 | 1104 | 491 | 276 | 177 | 36 |
| 0.09 | 15537 | 3884 | 1726 | 971 | 432 | 243 | 155 | 32 |
| 0.10 | 13830 | 3457 | 1537 | 864 | 384 | 216 | 138 | 29 |
| 0.15 | 8708 | 2177 | 968 | 544 | 242 | 136 | 87 | 19 |
| 0.20 | 6147 | 1537 | 683 | 384 | 171 | 96 | 61 | 14 |
| 0.25 | 4610 | 1152 | 512 | 288 | 128 | 72 | 46 | 11 |
| 0.30 | 3585 | 896 | 398 | 224 | 100 | 56 | 36 | 9 |
| 0.40 | 2305 | 576 | 256 | 144 | 64 | 36 | 23 | 6 |
| 0.50 | 1537 | 384 | 171 | 96 | 43 | 24 | 15 | 5 |

SOURCE: H. R. Bernard and P. D. Killworth, ''Sampling in Time Allocation Research,'' *Ethnology*, Vol. 32, p. 211. 

Here's how to read the table. Suppose people spend about 5% of their time eating. This is shown in the first column as a frequency, *f*, of 0.05. If you want to estimate the frequency of the activity to within 20%, look across to the column in the center part of table 14.2 under 0.20. If you have 1,825 observations, and your data say that people eat 5% of the time, then you can safely say that the true percentage of time spent eating is between 4% and 6%. (Twenty percent of 5% is 1%; 5%, plus or minus 1%, is 4%–6%. For the formula used to derive the numbers in table 14.2, see Bernard and Killworth 1993.)

Suppose you do a study of the daily activities of families in a community and your data show that men eat 4% of the time and women eat 6% of the time. If you have 300 observations, then the error bounds of the two estimates overlap considerably (about 0.02–0.06 for the men and 0.04–0.08 for the women).

You need about 1,800 observations to tell whether 0.06 is really bigger than 0.04 comparing across groups. It's the same for other activities: If women are seen at leisure 20% of their time and caring for children 25% of their time, then, as table 14.2 shows, you need 1,066 observations to tell if women really spend more time caring for children than they do at leisure.

Oboler had 1,500 observations. It is clear from table 14.2 that her findings about men's and women's leisure and work time are not accidents. An activity seen in a sample of just 256 observations to occur 40% of the time can be estimated actually to occur between 40%, plus or minus 15% of 40%, or between 34% and 46%. Since men are seen working 38% of the time and about half of Oboler's 1,500 observations were of men, her finding is solid.

### Nighttime Sampling

Most time allocation studies are done during the daylight hours, between 6 A.M. and 7–8 P.M. In Johnson's case, this was explicitly because "travel after dark is hazardous, and because visiting at night is not encouraged by the Machiguenga" (Johnson 1975:303).

Of course, we know that life doesn't stop when the sun goes down. Kenneth Good spent 13 years with the Yanomami in the Venezuelan Amazon. The Yanomami, Good reports, sit around at night and plan the next day's hunt—not in whispers, but in full volume. When the mood strikes, a Yanomami man might sit in his hammock and give a speech, to everyone within earshot. ''A Yanomami night,'' says Good, ''was like another day. . . . Like the Yanomami, I'd spend eleven hours in my hammock at night to get seven or eight hours of actual sleep'' (Good 1991:41–42).

Richard Scaglion (1986) did nighttime spot observations in his study of the Abelam in New Guinea. In 1983, when Scaglion did his study, there were 350 people in the village, living in 100 households. Scaglion randomly selected 2 households each day, and visited them at randomly selected times, throughout the day *and night*.

Scaglion didn't get much sleep during the month that he did this work, but his findings were worth the sacrifice. He coded his 153 observations into 13 categories of activities: sleeping, gardening, idle, cooking and food preparation, ritual, visiting, eating, hunting, construction, personal hygiene, child care, cleansing and washing, and craftwork. Only 74% of Abelam activities *during nighttime hours* were coded as ''sleeping.'' Seven of the nine observations that he coded as ''ritual'' occurred after dark. Half of all observations coded as ''hunting'' occurred at night, and six out of eight observations coded as ''visiting'' were nocturnal.

Had he done his TA study only during the day, Scaglion would have overestimated the amount of time that Abelam people spend gardening by about a fourth. His data show that gardening takes up about 26% of the Abelam's daylight hours, but only 20% of their total waking time in each 24-hour period.

It may not always be possible to conduct TA studies at night. Johnson, you'll remember, made a point of the fact that the Machiguenga discourage nighttime visiting. Scaglion, on the other hand, worked among a people who ''go visiting at unusual hours, even when their prospective host is likely to be sleeping.''

Scaglion, in fact, rather enjoyed showing up at people's houses during odd hours in 1983. He had done his doctoral research in the same village in 1974–75. In those days, he says, ''I was still quite a novelty. . . . I was frequently awakened by hearing '*Minoa, mine kwak*?' ('Hey, you, are you sleeping?'). This study allowed me to return old favors by visiting people in the late night hours to be sure *they* were sleeping'' (Scaglion 1986:539) (**Further Reading:** time allocation studies).

### Coding and Recording Time Allocation Data

Sampling is one of two problems in TA research. The other is measurement. How do we know that when Oboler recorded that someone was ''working,'' we would have recorded the same thing? If you were with Johnson when he recorded that someone was engaged in ''hygiene behavior,'' would you have agreed with his assessment? Every time? You see the problem.

It gets even more thorny. Suppose you work out a coding scheme that everyone agrees with. And suppose you train other observers to see just what you see. (Rogoff [1978] achieved a phenomenal 98% interobserver agreement in her study of 9 year olds in Guatemala.) Or, if you are doing the research all by yourself, suppose you are absolutely consistent in recording behaviors (i.e., you never code someone lying in a hammock as sleeping when they're just lounging around awake).

Even if all these reliability problems are taken care of, what about observation validity? What do you do, for example, when you see people engaged in multiple behaviors? A woman might be holding a baby and stirring a pot at the same time. Do you code her as engaged in child care or in cooking? (Gross 1984:542). If someone saw that you were lying down reading and you were studying for an exam, should they record that you were working or relaxing?

Do you record all behaviors? Do you mark one behavior as primary? This last question has important implications for data analysis. There are only so many minutes in a day, and the percentage of people's time that they allocate to activities has to add up to 100%. If you code multiple activities as equally important, then there will be more than 100% of the day accounted for. Most TA researchers use their intuition, based on participant observation, to decide which of the multiple simultaneous activities they witness to record as the primary one and which as secondary.

The best solution is to record *all* possible behaviors you observe in the order of their primacy, according to your best judgment at the time of observation. Use a check sheet to record behaviors. Use a separate check sheet for each observation you make. This can mean printing up 1,000 sheets for a TA study, and hauling them home later. You can save yourself a lot of work by using a hand-held computer with the equivalent of an electronic check sheet installed that lets you watch behavior and code it on the spot (see appendix E) (box 14.4).

**BOX 14.4**

### ON BEING PROPERLY PARANOID ABOUT LOSING DATA

Saving work is one of two good reasons to use hand-held computers to record observational data. The other good reason to give up paper check sheets if you can is that, eventually, you have to type in the data anyway, transferring them from paper to computer so you can run statistical analyses. Each time you record behavioral data—as you watch it and as you transfer it from paper to computer—there is a chance of error. That's just how it is. So eliminating one step in the data management process cuts down on errors in data.

If you must hand-code your original observations, enter the data into a laptop while you're still in the field as a precaution against loss of the original data sheets. Be paranoid about data. Those horror stories you've heard about lost data? They're true. (Read M. N. Srinivas's account [1979:xiii] of how he lost all three copies of his field notes, compiled over a period of 18 years, in a fire at Stanford.)

## EXPERIENCE SAMPLING

In **experience sampling** (ES), people respond at random times during a day or a week to questions about what they're doing, or who they're with, or what they're feeling at the moment. Some researchers ask informants to carry around a beeper and, when the beeper goes off, to jot an entry into a diary or talk about their actions, feelings, and surroundings into a small digital recorder. Some researchers ask informants to respond to a telephone interviewer (Kubey et al. 1996), and some researchers ask informants to fill out a form on an Internet-enabled cell phone or PDA (Foo et al. 2009; Wenze et al. 2007).

ES offers two big advantages. First, it combines the power of random spot checks with the relative ease of having people report on their own behavior. Csikszentmihalyi and Larson (1987) demonstrated the reliability of ES in a number of studies. Validity is another matter, but when people record or talk about what they're doing and how they're feeling on the spot, this should lessen the inherent inaccuracy of recall data. The second advantage is that, working on your own, you can only be in one place at a time, but with those beepers or cell phones, you can collect spot-observation data from lots of people at once.

In anthropology, Garry Chick used ES in his study (1994) of a small machine-tool company in Pennsylvania. Chick wanted to test Marx's theory that automation would eliminate the need for skilled labor by turning complex tasks into a series of routine steps that could be performed by drones. Drones should find their work boring, unsatisfying, and unimportant.

There were 11 machinists, all men, in the small shop. Four of them worked on traditional, manually controlled lathes in turning out machine tools. Three worked only on a newly installed, computer-controlled lathe (you program it to do a job and then you more-or-less stand by while it executes your instructions). And four worked on both types of lathes, depending on the job.

Each man was beeped 30 times—six times a day over a 5-day week—and at each beep, he stopped what he was doing within 10 minutes and filled out a 2-minute questionnaire. (Each man kept a spiral-bound booklet of questionnaires handy in the shop.) There should have been 330 questionnaires (11 men filling out 30 each), but one man was beeped while he was in the bathroom and one was beeped while he was filling out a questionnaire (having been beeped 2 minutes before). That's what happens when you collect data at random points, but random means random and no tinkering with it.

For each of the experience samples, workers indicated what they were doing. If they were operating a machine (or more than one), they indicated which one(s). Then they answered 14 attitude questions. Here they are:

On a scale from 1–10, where 1 means "definitely not" and 10 means "definitely yes," what I was doing:

1. was enjoyable.
2. was interesting.
3. was complex/technical.
4. was fun.
5. was under my control.
6. was monotonous.
7. was machine paced.
8. was tricky.
9. held my attention.

On a scale from 1–10, where 1 means "definitely not" and 10 means "definitely yes," at the time I was signaled, I was:

10. pressed for time.
11. working on my own.
12. thinking about things other than work.
13. doing something that I felt was important.
14. doing something that required a lot of skill.

The results were mixed. The men found working on the computer-controlled machines more interesting, more satisfying, and more important than working on the manual machines, and they found programming the computer-controlled machines even more interesting than running them. But these machinists also felt more in control when they worked on manual lathes.

If you use self-administered questionnaires, you need literate informants to do experience sampling. But, as Chick says, "with more and more anthropological research in modern and modernizing societies, experience sampling can be a valuable addition to the anthropologist's tool kit" (1994:6). And if you give people little voice recorders, you may be able to use the ES method with nonliterate populations (**Further Reading:** experience sampling).

## COMBINING CONTINUOUS MONITORING AND SPOT SAMPLING

The difference between CM and spot sampling is analogous to the difference between ethnography and survey research. With ethnography, you get information about process; with survey research, you get data that let you estimate parameters for a population. Spot sampling is used in TA research precisely because the goal is to estimate parameters—like how much time, on average, women spend cooking, or men spend throwing pots, or children spend laid up ill at home. If you want to know the ingredients of *mafongo* (a dish native to Puerto Rico) and the order in which they are added, you have to watch continuously as someone makes it. (Making it yourself, as part of participant observation, produces embodied knowledge, yet a third kind of information.)

Robin O'Brian (1998) combined CM and spot sampling in her study of Mayan craftswomen in Chiapas, Mexico. The women sold their crafts to tourists at a local market. Several times a week, O'Brian went through the market (entering from a randomly selected spot each time) and coded what every woman craft seller was doing, using a check sheet adapted from A. Johnson et al.'s (1987) standardized time allocation activity codes.

O'Brian also did continuous monitoring for 3 hours of 15 women, and these two kinds of data produced more information than either kind alone. Her aggregate, spot-sampling data showed that the women spent 82% of their time waiting for tourists to buy something. The women weren't just sitting around, though. They spent 17% of their waiting time producing more crafts (doing macramé or embroidery or hand-spinning wool) and another 17% eating, cleaning children, or nursing. The CM data showed that women were interrupted in their productive activities 36% of their waiting time (that is, 36% of 82%, or 30% of their entire day) and that the interruptions were as likely to be responding to their children's needs as they were to be selling to tourists.

## PLUSES AND MINUSES OF DIRECT OBSERVATION

On balance, direct observation provides much more accurate results about behavior than do reports of behavior. Ricci et al. (1995) studied 40 people in Kalama, a peri-urban village about 15 miles north of Cairo. One day, a set of trained observers watched the 40 participants for 2.5 hours. The observers noted whether people were engaged in any of 17 activities and how long each person spent at each activity.

The next day, the participants were asked to recall, sequentially, what they had done the entire day before, and how long they had spent at each activity. The interviewers did not mention any of the 17 activities, but they tried to improve respondent recall by asking about activities before and after locally significant time markers, like call to prayer. Ten of the 40 were toddlers, so Ricci et al. focused on the recall data of the 24 adults and the six school-age children.

Ricci et al. were very forgiving in their analysis. Informants were scored as being correct if they could recall an activity at all and say correctly whether it had been done in the morning or afternoon observation period (9:00–11:30 A.M. or 12:30–3:00 P.M.). Ricci et al. scored only errors of omission (leaving out activities) and threw out errors of commission (inventing activities that had not been observed).

And informants *still* got it wrong—a lot. Across men and women, across agricultural and nonagricultural households, informants got it wrong, on average, 56% of the time. Five of the 6 women who had been observed breast-feeding failed to report that activity the next day; 13 of the 15 women who had been observed washing clothes failed to report that activity the next day.

If you want to know whether, say, caring for animals happens more often than, say, gathering fuel, self-reports might be enough. But if you want to know *how often* those behaviors actually occur, then nothing short of direct observation will do.

I don't want to give the impression, however, that direct observation data are automatically accurate. Lots of things can clobber the accuracy of directly observed behavior. Observers may be biased by their own expectations of what they are looking for or by expectations about the behavior of women or men or any ethnic group (Kent et al. 1977; Repp et al. 1988).

You may feel awkward about walking around with a clipboard (and perhaps a stopwatch) and writing down what people are doing—or with beeping people and asking them to interrupt what they're doing to help you get some data. This is a reasonable concern, and direct observation is not for everyone. It's not a detached method, like sending out questionnaires and waiting for data to be delivered to your doorstep.

It is not a fun method, either. Hanging out, participating in normal daily activities with people, and writing up field notes at night is more enjoyable than monitoring and recording what people are doing.

But many fieldworkers find that direct observation allows them to address issues that are not easily studied by any other method. Grace Marquis (1990) studied a shantytown in Lima, Peru. Children in households that kept chickens were at higher risk for getting diarrhea than were other children. The chickens left feces in the homes, and the feces contained an organism that causes diarrhea. Continuous monitoring showed that children touched the chicken droppings and, inevitably, touched their mouths with their hands. It was hard, tedious work, but the payoff was serious.

Direct observation is time consuming, but random spot-checking of behavior is a very cost effective and productive way to use *some* of your time in any field project. When you're studying a group that has clear boundaries (a village, a hospital, a school), you can get very fine-grained data about people's behavior from a TA study, based on random spot checks. More importantly, as you can see from table 14.2, with proper sampling you can generalize to large populations (whole school districts, an entire aircraft manufacturing plant, even cities) from spot checks of behavior, in ways that no other method allows.

You may be concerned that a strictly observational approach to gathering data about human behavior fails to capture the *meaning* of data for the actors. This, too, is a legitimate concern. A classic example is Geertz's (1973:6–7) observation that a wink can be the result of getting a speck of dust in your eye or a conscious act of conspiracy. And that's just a wink. People can engage in any of thousands of behaviors (skipping a class, wearing a tie, having their navel pierced . . .) for many, many different reasons. Knowing the meaning of behavior to others is essential to understanding it ourselves.

On the other hand, one of our most important goals in science is to constantly challenge our own ideas about what things mean. That's how theories develop, are knocked down, and gain in their power to explain things. Why shouldn't we also challenge the theories—the explanations—that the people we study give us for their own behavior?

Ask people who are coming out of a church, for example, why they just spent 2 hours there. Some common responses include "to worship God," "to be a better person," "to teach our children good values." Hardly anyone says "to dress up and look good in front of other people," "to meet potential golf partners for later this Sunday afternoon," or "to maximize my ability to meet potential mates whose ethnic and social backgrounds are compatible with my own." Yet, we know that these last three reasons are what *some* people would say if they thought others wouldn't disapprove.

Finally, you may have some qualms about the ethics of obtrusive observation. It cannot be said too often that *every single data collection act* in the field has an ethical component, and a fieldworker is obliged every single time to think through the ethical implications of data collection acts. Personally, I have less difficulty with the potential ethical problems of obtrusive, reactive observation than I do with any other data collection method, including participant observation. In obtrusive observation, people actually *see* you (or a camera) taking down their behavior, and they can ask you to stop. Nothing is hidden.

In participant observation, we try to put people at ease, make them forget we're really listening hard to what they're telling us, and get them to "open up." We ask people to take us into their confidence, and we are handed the responsibility for not abusing that confidence.

But the method that presents the *most* ethical problems is unobtrusive, *non*reactive direct observation.

## UNOBTRUSIVE OBSERVATION

**Disguised field observation** is the ultimate in participant observation—you join, or pretend to join, some group and secretly record data about people in the group.

In 1960, John H. Griffin, a white journalist went through some drug treatment to temporarily turn his skin black. He traveled the southern United States for about a month, taking notes on how he was treated. His book, *Black Like Me* (1961) was a real shocker. It galvanized a lot of support by Whites in the North for the then fledgling Civil Rights movement. Clearly, Griffin engaged in premeditated deception in gathering the data for his book. But Griffin was a journalist; scientists don't deceive their informants, right?

### Pseudopatients and Simulated Clients

Wrong. Samuel Sarkodie, an M.A. student in medical sociology at the University of Legon, in Ghana, spent 3 days in a rural hospital in 1994 as a pseudopatient with a false case of malaria. The hospital staff were in on the study—they had been recruited by Sarkodie's supervisor, Sjaak van der Geest, a Dutch anthropologist who works in Ghana—and Sarkodie wrote a detailed report for the hospital. Presumably, the report was helpful, but Sarkodie's fellow patients were duped (van der Geest and Sarkodie 1998).

Twenty years earlier, David Rosenhan recruited seven confederates who, like him, checked themselves into mental hospitals and took surreptitious notes about how they were treated. They gave false names and occupations (they couldn't very well mention their real occupations since three of them were psychologists and one was a psychiatrist), and reported hearing voices. One was diagnosed as manic-depressive, and the rest as schizophrenics, and all were admitted for treatment.

This was tough work. The pseudopatients were not allowed to divulge what they were up to just because they were tired of (or exasperated with) the experiment. The only way out was to be diagnosed by the hospital staff as ready for release. It took between 1 week and 7 weeks of confinement to achieve this, and when they were released, the pseudopatients were all diagnosed with ''schizophrenia in remission'' or as ''asymptomatic'' or as ''improved'' (Rosenhan 1973, 1975).

Rosenhan's field experiment made clear the power of labeling: Once you are diagnosed as insane, people treat you as insane. Period. Some of the genuine inmates at the hospitals saw through the charade, but none of the staff ever did (box 14.5).

**BOX 14.5**

### ARE THESE PSEUDOPATIENT STUDIES ETHICAL?

The simulated client method has been used in dozens of studies to evaluate the performance of physicians, pharmacists, family-planning clinics, and other health care providers in developing nations. (See Madden et al. [1997] for a review of these studies.) And fake clients—men and women, black, white, and Hispanic—are sent out by U.S. government agencies regularly to apply for jobs, to rent apartments, or to buy homes and to uncover discrimination (Sharpe 1998). The U.S. Supreme Court has ruled that this practice is legal in the pursuit of fair housing (Ayres 1991:823), and the Equal Employment Opportunity Commission uses data from these field experiments to sue offending businesses.

People across the political spectrum have quite different ideas about whether this is just a dose of the same medicine that offenders dish out (which seems fair), or entrapment (which seems foul). Does this mean that ethics are simply a matter of political orientation and opinion? In the abstract, most people answer this question with a strong ''no.'' When things get concrete—when the fortunes and reputations of real people are at stake—the answer becomes less clear (van den Borne 2007).

But if you think deceiving landlords or Realtors or the staff of mental hospitals is something, read on (**Further Reading:** pseudopatients and simulated clients).

**The Tearoom Trade Study**

Without telling people that he was studying them, Laud Humphreys (1975) observed hundreds of homosexual acts among men in St. Louis, Missouri. Humphreys's study produced very important results. The men involved in this tearoom trade, as it is called, came from all walks of life, and many were married and living otherwise straight lives. Humphreys made it clear that he did not engage in homosexual acts himself, but played the role of the ''watch queen,'' or lookout, warning his informants when someone approached the restroom. This deception and unobtrusive observation, however, did not cause the storm of criticism that accompanied the first publication of Humphreys's work in 1970.

That was caused by Humphreys having taken his research a step further. He jotted down the license plate numbers of the men who used the restroom for quick, impersonal sex and got their names and addresses from motor vehicle records. He waited a year after doing his observational work, and then, on the pretext that they had been randomly selected for inclusion in a general health survey, he interviewed 100 of his research subjects in their homes.

Humphreys was careful to change his car, his hairstyle, and his dress. According to him, his informants did not recognize him as the man who had once played watch queen for them in public toilets. *This* is what made Humphreys's research the focus of another debate, which is still going on, about the ethics of nonreactive field observation.

Five years after the initial study was published, Humphreys himself said that he had made a mistake. He had endangered the social, emotional, and economic lives of people he studied. Had his files been subpoenaed, he could not have claimed immunity. He decided at the time that he would go to jail rather than hurt his informants (Humphreys 1975).

Humphreys was an ordained Episcopal priest who had held a parish for more than a decade before going to graduate school. He was active in the Civil Rights movement in the early 1960s and spent time in jail for committing crimes of conscience. His credentials as an ethical person, conscious of his responsibilities to others, were in good order. Everyone associated with him agreed that Humphreys was totally committed to protecting his informants.

But listen to what Arlene Kaplan Daniels had to say about all this, in a letter to Myron Glazer, a sociologist and ethnographer:

> In my opinion, no one in the society deserves to be trusted with hot, incriminating data. Let me repeat, *no one*. . . . We should not have to rely on the individual strength of conscience which may be required. Psychiatrists, for example, are notorious gossipers [about their patients]. . . . O. K., so they mainly just tell one another. But they *sometimes* tell wives, people at parties, you and me. [Daniels had done participant observation research on psychiatrists.] And few of them would hold up under systematic pressure from government or whatever to get them to tell. . . . The issue is not that a few brave souls *do* resist. The issue is rather what to do about the few who will not. . . . There is *nothing* in our training—any more than in the training of psychiatrists, no matter what they say—to prepare us to take up these burdens. (quoted in Glazer 1975:219–20; emphasis in original)

Researchers who conduct the kinds of studies that Humphreys did invoke several arguments to justify the use of deception.

1. It is impossible to study such things as homosexual encounters in public restrooms in any other way.
2. Disguised field observation is a technique that is available only to researchers who are physically and linguistically indistinguishable from the people they are studying. To use this technique, you must be a member of the larger culture. There is, therefore, no real ethical question involved, other than whether you, as an individual, feel comfortable doing this kind of research.
3. Public places, like restrooms, are, simply, public. The counterargument is that people have a right to expect that their behavior in public toilets will not be recorded, period. (Koocher 1977)

Sechrest and Phillips (1979) take a middle ground. They say that ''public behavior should be observable by any means that protect what might be called 'assumed' privacy, the privacy that one might expect from being at a distance from others or of being screened from usual views'' (p. 14). Casual observation is fine, but the use of telescopes, listening devices, or peepholes would be unethical.

My own position is that the decision to use deception is up to you, provided that the *risks of detection are your own risks and no one else's*. When Jack Weatherford (1986) took a job as manager of a porn shop in Washington, DC, the people who came to the store to watch the movies or connect with prostitutes didn't know they were being studied by a participant observer, but neither were they in any danger that their identities would be divulged. And similarly, when Wendy Chapkis became a licensed massage therapist and became a participant observer in her secret research on prostitution (1997), she assumed risks, but the risks were hers. If detection risks harm to others, then don't even consider disguised participant observation. Recognize, too, that it may not be possible to foresee the potential harm that you might do using disguised observation. This is what leads scholars like Kai Erikson (1967, 1996) to the conclusion that research that requires deception is never justified (**Further Reading:** deception in participant observation).

## GRADES OF DECEPTION

But is all deception equally deceitful? Aren't there **grades of deception**? In the 1960s, Edward Hall and others (Hall 1963, 1966; Watson and Graves 1966) showed how people in different cultures use different ''body language'' to communicate—that is, they stand at different angles to one another, or at different distances when engaging in serious versus casual conversation. Hall called this different use of space **proxemics**. He noted that people learn this proxemic behavior as part of their early cultural learning and he hypothesized that subcultural variations in spatial orientation often leads to breakdowns in communication, isolation of minorities, and so on.

This seminal observation by an anthropologist set off a flurry of research that continues to this day (for a review, see Farnell 1999). Early on, Aiello and Jones (1971) studied the proxemic behavior of middle-class white and lower-class Puerto Rican and black schoolchildren. They trained a group of elementary schoolteachers to observe and code the distance and orientation of pairs of children to one another during recess periods (**Further Reading:** proxemics).

Sure enough, there were clear cultural and gender differences. White children stand much farther apart in ordinary interaction than do either black or Puerto Rican children. The point here is that the teachers were natural participants in the system. The researchers trained these natural participants to be observers to cut out any reactivity that outsiders might have caused in doing the observation (box 14.6).

Levine (1997) used casual, unobtrusive observation to study the walking speed of people in different size cities, and Rotton et al. (1990) tested whether people walk faster in a climate-controlled mall or in an open-air shopping. Contrary to popular wisdom, heat didn't slow down urban shoppers of either sex. And Sykes et al. (1993) sat unobtrusively in bars, counting the number of drinks people consumed. Confirming popular wisdom, people drink faster and spend less time in bars when they are in groups of two or more than when they're alone.

I don't consider these field studies of shoppers, children, pedestrians, and drinkers in bars to be unethical. The people being studied were observed in the course of their ordinary activities, out in the open, in truly public places. Despite making unobtrusive observations or taking surreptitious pictures, the deception involved was passive—it didn't involve ''taking in'' the subjects of the research, making them believe one thing to get them to do another. I don't think that any real invasion of privacy occurred.

**BOX 14.6**

### LOWERING REACTIVITY IN DIRECT OBSERVATION RESEARCH

For his master's degree, Mark House studied shoppers in Nizhny Novgorod, a city on the Volga River about 400 miles east of Moscow in the Volga region. As he followed shoppers around, House held his little tape recorder to his ear, as if he were talking on a cell phone, and dictated notes. Shawn Scherer (1974) studied pairs of children in a schoolyard in Toronto. He used only lower-class black and lower-class white children in his study, to control for socioeconomic effects. Scherer adapted techniques from photogrammetry (making surveys by using photographs). He mounted a camera in a park adjacent to the schoolyard. Using a telephoto lens, he took unobtrusive shots of pairs of children who were at least 30 meters away.

This got rid of the reactivity problem. Then Scherer devised a clever way to measure the average distance between two children and did his analysis on the quantitative data. Scherer found no significant differences in the distance between pairs of white or black children.

**The Micturition Study**

You can't say that about the work of Middlemist et al. (1976). They wanted to measure the length of time it takes for men to begin urinating, how long men continue to urinate, and whether these things are affected by how close men stand to each other in public toilets. (*Why* they wanted to know these things is another story.)

At first, the researchers pretended to be combing their hair at the sink in a public toilet at a university. They tracked the time between the sound of a fly being unzipped and urine hitting the water in the urinal as the time for onset, then they noted how long it took for the sound of urine to stop hitting the water in the urinal and counted this as the duration of each event. They noted whether subjects were standing alone, next to someone, or one or two urinals away from someone.

In general, the closer the man being watched stood to another man, the longer it took him to begin urinating and the shorter the duration of the event. This confirmed laboratory research showing that social stress inhibits relaxation of the urethral sphincter in men, thus inhibiting flow of urine.

Middlemist et al. decided to control the independent variable—how far away another man was from each subject. They placed ''BEING CLEANED'' signs on some urinals and forced unsuspecting men to use a particular urinal in a public toilet. Then a confederate stood next to the subject, or one urinal away, or did not appear at all. The observer hid in a toilet stall next to the urinals and made the measurements. The problem was, the observer couldn't hear flies unzipping and urine hitting the water from inside the stall—so the researchers used a periscopic prism, trained on the area of interest, to make the observations directly.

Personally, I doubt that many people would have objected to the study if Middlemist and his colleagues had just lurked in the restroom and done simple, unobtrusive observation. But when they contrived to make men urinate in a specific place, when they contrived to manipulate the dependent variable (urination time), and, above all, when they got that periscope into the act, that changed matters. This is a clear case of invasion of privacy by researchers, in my view.

In a severe critique of the research, Koocher (1977:120) said that "at the very least, the design seems laughable and trivial." Middlemist et al. (1977:123) defended themselves, saying that "we believe . . . that the pilot observation and the experiment together constitute an example of well-controlled field research, adequate to test the null hypothesis that closeness has no effect" on the duration of urination among men in public restrooms. Actually, Middlemist et al.'s study *design* was anything but trivial. In fact, it was quite elegant. And the results of their research have been cited many times in articles on the stress of crowding—like why many people prefer to stand in commuter trains rather than sit in middle seats between other passengers (Evans and Wener 2007)—and on paruresis, an anxiety disorder that inhibits urination in the presence of, or anticipated presence of others (Boschen 2008). Whether the research was ethical is another matter.

### Passive Deception

**Passive deception** involves no experimental manipulation of informants to get them to act in certain ways. Humphreys's first, strictly observational study (not the one where he used a pretext to interview people in their homes) involved passive deception. He made his observations in public places where he had every right to be in the first place. He took no names down, and there were no data that could be traced to any particular individual. Humphreys observed felonies, and that makes the case more complex. But in my opinion, he had the right to observe others in public places, irrespective of whether those observed believed that they would or would not be observed. What he did with his observations—following people up by tracking them through their license plates—is, like Middlemist et al.'s periscope, another matter.

Anthropologists use passive deception all the time. I have spent hours pretending to be a shopper in department stores and have observed mothers who are disciplining their children. I have played the role of a strolling tourist on Mexican beaches (an easy role to play since that was exactly what I was) and recorded how American and Mexican families occupied beach space. I have surreptitiously clocked the time it takes for people who were walking along the streets of Athens (Greece), New York City, Gainesville (Florida), and Ixmiquilpan (Mexico) to cover 10 meters of sidewalk at various times of the day. I have stood in crowded outdoor bazaars in Mexico, watching and recording differences between Indians and non-Indians in the amount and kinds of produce purchased.

I have never felt the slightest ethical qualm about having made these observations. In my opinion, passive deception is ethically aseptic. Ultimately, however, the responsibility for the choice of method, and for the practical, human consequences of using a particular method, rests with you, the individual researcher. You can't foist off that responsibility on "the profession," or on some "code of ethics."

Are you disturbed by the fact that Humphreys did his research at all, or only by the fact that he came close to compromising his informants? As you answer that question for yourself, you'll have a better idea of where *you* stand on the issue of disguised field observation. (For more on the ethics of deception, see **Further Reading**, chapter 4.)

## BEHAVIOR TRACE STUDIES

Think of trace studies as behavioral archeology. Do people in different cultures really have a different sense of time? Levine and Bartlett (1984) went to 12 cities in six countries and noted the time on 15 randomly chosen bank clocks in each city. Then they measured the difference between the time shown on the clocks and the time reported by the local telephone company in each city. The most accurate public clocks were Japanese—off by an average of just 34 seconds. U.S. clocks were next (off by an average of 54 seconds), followed by the clocks in Taiwan, England, and Italy (71 sec., 72 sec., and 90 sec., respectively). Indonesia came in last, at 189 sec.

Here you have hard, archeological evidence of clock-setting behavior across six countries. Real people had set those 15 clocks in each city, and real people were responsible for making sure that the clocks were adjusted from time to time. Levine and Bartlett looked at whether differences in the average deviation of the clocks from the real time predicted differences in the rate of heart disease.

They don't. The country with the lowest rate of heart disease, Japan, has the most accurate clocks and the fastest overall pace of life (as measured by several other indicators). Apparently, according to Levine and Bartlett, it's possible in some cultures to be hard *working* without being hard *driving*.

Sechrest and Flores (1969) recorded and analyzed bathroom graffiti in a sample of men's public toilets in Manilla and Chicago. They wanted to examine attitudes toward sexuality in the two cultures. The results were striking. There was no difference in the percentage of graffiti in the two cities that dealt with heterosexual themes. But fully 42% of the Chicago graffiti dealt with homosexuality, whereas only 2% of the Manilla graffiti did, showing a clear difference in the two cultures regarding level of concern with homosexuality.

Gould and Potter (1984) did a survey of used-up (not smashed-up) automobiles in five Providence, Rhode Island, junkyards. They calculated that the average use-life of American-made cars is 10.56 years, irrespective of how many times cars change hands. This is a good deal longer than most Americans would guess. Gould also compared use-life against initial cost and found that paying more for a car doesn't affect how long it will last. Interesting and useful findings.

In their classic book on *Unobtrusive Measures*, Webb et al. (1966) identified a class of measures based on erosion. Administrators of Chicago's Museum of Science and Industry had found that the vinyl tiles around an exhibit showing live, hatching chicks needed to be replaced about every 6 weeks. The tiles around other exhibits lasted for years without having to be replaced. Webb et al. (p. 37) suggested that this erosion measure (the rate of wear on vinyl tiles) might be a proxy for a direct measure of the popularity of exhibits. The faster the tiles wear out, the more popular the exhibit (box 14.7).

### The Garbage Project

The Garbage Project was founded in 1973 by archeologist William Rathje at the University of Arizona. For over 25 years, Rathje and his associates studied consumer behavior patterns in Tucson, Arizona, by analyzing the garbage from a representative sample of residents. It was a great effort at applying trace measures.

In 1988, about 6,000 residents of Tucson were sent flyers, explaining that they were selected to be part of a study of recycling behavior. Their garbage would be studied, the flyer explained, and confidentiality was assured, but if they didn't want to be part of the study, residents could send in a card and they would be removed from the list. About 200 people returned the cards and opted out of the study (Wilson Hughes, personal communication). (And see Hughes [1984] for a detailed review of the methodology of the Garbage Project.)

By studying the detritus of ordinary people, researchers on the Garbage Project learned

**BOX 14.7**

### WEIGHING THE EVIDENCE

Dean Archer and Lynn Erlich (1985) had a hypothesis that sensational crimes (with a lot of press coverage) result in increased sales of handguns. The police would not allow them to see the handgun applications, so they asked a member of the police staff to put the permits into envelopes, by month, for 3 months before and 3 months after a particular sensational crime. Then they weighed the envelopes and converted the weight to handgun applications. To do this, they got a chunk of blank applications and found out how many applications there were per ounce.

The technique is very reliable. The correlation between the estimates of researchers and the actual weights of the envelopes was .99, and in a controlled experiment, researchers were able to tell the difference of just one sheet of paper in 15 out of 18 tries. Real data can be messy, though. Lots of handgun applications have addenda attached, for example. Still, the correlation between researchers' estimates and the true number of handgun applications across 6 months was .94.

As Archer and Erlich suggest, the weight method can be used to study confidential records when you want to know only aggregate outcomes—about things like drunk driving arrests, the influx of psychiatric patients to a clinic, the number of grievance filings in a company, the number of abortion referrals, and the number of complaints against agencies—and don't need data about individuals.

interesting things about food consumption and waste among Americans. Squash is the favored baby food among Hispanics in the United States, and 35% of all food from chicken take-out restaurants is thrown away (Rathje 1992). You can accurately estimate the population of an area by weighing only the plastic trash. Children, it turns out, generate as much plastic trash as adults do (Edmondson 1988).

Early in the Garbage Project, researchers expected that people would not waste much beef during a shortage, but exactly the opposite happened in 1973. Two things were shown to be responsible for this finding. First, as the shortage took hold, the price of beef rose, and people started buying cheaper cuts. Some residents did not know how to prepare those cuts properly, and this created more waste; others found that they didn't like the cheaper cuts and threw out more than they usually would have; and cheaper cuts have more waste fat to throw out to begin with. Second, as the price continued to rise, people started buying greater quantities of beef, perhaps as a hedge against further price hikes. Inevitably, some of the increased purchases spoiled from lack of proper storage (Rathje 1984:17).

Rathje found the same pattern of consumer behavior during the sugar shortage of 1975. He reasoned that whenever people changed their food-buying and -consuming habits drastically, there would be at least a short-term increase in food loss. Conversely, when people use foods and ingredients that are familiar to them they waste less in both preparation and consumption.

This led Rathje to compare the food loss rate among Mexican Americans and Anglos in Tucson and Milwaukee. "The final results of Mexican-American cooking," Rathje said, "can be extremely varied—chimichangas, burros, enchiladas, tacos, and more—but the basic set of ingredients are very few compared to standard Anglo fare. Thus, Mexican-American households should throw out less food than Anglo households" (Rathje 1984:17–18). In fact, this is exactly what Rathje found in both Tucson and Milwaukee.

Beside Tucson and Milwaukee, studies of fresh household garbage have been done in New Orleans, Marin County (California), Mexico City, and Sydney (Australia).

**Pros and Cons of Trace Studies**

The most important advantage of trace studies is that they are nonreactive, so long as the people you are studying are kept in the dark about what you are doing. What happens when people are told that their garbage is being monitored? Ritenbaugh and Harrison (1984) compared data from an experimental group (people who were told that their garbage was being monitored) and a control group (people who were not told). There was no difference in the refuse disposal behavior of the experimental and control groups—with one important exception. The number of empty bottles of alcoholic drinks that showed up was significantly lower when people knew that their garbage was being monitored.

Where did the extra bottles go? Buried in the backyard? Stuffed in the trash cans of neighbors who were not in the sample? It remains a mystery.

In addition to being nonreactive, behavioral trace studies yield enormous amounts of data that can be standardized, quantified, and compared across groups and over time (Rathje 1979). Moreover, traces reflect many behaviors more accurately than informant reports of those behaviors.

In 1986, as part of a contract with the Heinz Corporation, Rathje and his colleagues asked women in the Tucson area if they had used any store-bought baby food in the past week. Uniformly, the Hispanic mothers insisted that they had not used any such product. You can guess the rest: They had as many baby-food jars in their garbage as did the Anglo households—and this, despite that fact that 45% of the Hispanic women in Tucson at the time were working outside the home (Rathje and Murphy 1992). Those women were caught in a bind: They didn't have time to prepare home-made baby food, but they couldn't admit this to a stranger asking questions.

Trace studies like the Garbage Project have plenty of problems, however. Early in the project, it became apparent that garbage disposals were going to be a serious problem. The researchers constructed a subsample of 32 households, some of which had disposals, some of which did not. They studied these 32 households for 5 weeks and developed a "garbage disposal correction factor" (Rathje 1984:16).

As the project went on, researchers learned that some families were recycling all their aluminum cans, and others were throwing theirs in the trash. This made it difficult to compare households regarding their consumption of soft drinks and beer. Some families had compost heaps that they used as fertilizer for their vegetable gardens. This distorted the refuse count for those families. Garbage Project researchers had to develop correction factors for all of these biases, too (see G. G. Harrison 1976).

As with much unobtrusive research, the Garbage Project raised some difficult ethical problems. To protect the privacy of the households in the study, no addresses or names of household members were recorded. All personal items, such as photographs and letters, were thrown out without being examined. The hundreds of student sorters who worked on the project signed pledges not to save anything from the refuse they examined. All the sampling, sorting, and data analysis procedures were approved by the Human Subjects Research Committee of the University of Arizona.

The Garbage Project received consistent coverage in the press, both nationally and locally in Tucson. In 1984, after 10 years of work, Hughes reported that ''no public concern over the issue of personal privacy has been expressed, and community response has been supportive'' (Hughes 1984:42). With proper safeguards, trace measures generate lots of useful data about human behavior.

## ARCHIVAL RESEARCH

One of the great advantages to doing archival research is that it is truly nonreactive. Whether you're studying archival records of births, migrations, visits to a hospital, or purchases of hybrid seed, people can't change their behavior after the fact. The original data might have been collected reactively, but that's one reason why historians demand such critical examination of sources.

Another advantage of doing what Caroline Brettell calls ''fieldwork in the archives'' (1998) is that you can study things using archival data that would be too politically ''hot'' to study any other way. And archival research is inexpensive. Be on the lookout for interesting archival materials: government reports, personal diaries or photo collections, industrial data, medical records, school records, wills, deeds, records of court cases, tax rolls, and land-holding records.

### Cultural Processes

Archival resources can be particularly useful in studying cultural processes through time. June Helm (1980) found that between 1829 and 1891, traders at the Hudson's Bay Company posts of the upper Mackenzie Delta had surveyed the Indians who traded at their stores. On the basis of those data, Helm concluded that, before 1850, the Indians of the area had practiced female infanticide. After 1850, missionaries were successful in stopping infanticide. Nancy Howell (1981), a demographer, subjected Helm's data to a sophisticated statistical analysis and corroborated Helm's conclusion.

Daniel Swan and Gregory Campbell (1989) studied the population records of 1877 to 1907 for the Osage reserve. They were able to show that from 1877 to 1887, the full bloods declined at 6.4% a year and the mixed bloods increased at 7.3% a year. This had great consequences for the Osage because the full bloods and mixed bloods had formed voting blocs on economic issues. In particular, the full bloods resisted turning the reserve land into private property. Whites who married into the tribe fraudulently claimed tribal mixed-blood status. The mixed bloods were in favor of the private property measures.

Using fashion magazines going back to 1844, Alfred Kroeber made eight separate measurement of women's clothing in the United States and France (Kroeber 1919). He measured things like the diameter of the skirt at the hem, the diameter of the waist, the depth of decolletage (measured from the mouth to the middle of the corsage edge in front), and so on. After analyzing the data, Kroeber claimed to have found ''an underlying pulsation in the width of civilized women's skirts, which is symmetrical and extends in its up and down beat over a full century; and an analogous rhythm in skirt length, but with a period of only about a third the duration'' (p. 257). Kroeber offered his finding as evidence for long-cycle behavior in civilization.

Allport and Hartman (1931) criticized Kroeber for having been insufficiently critical of his sources. They found, for example, that the range in width of skirts for one year, 1886, was greater than the range Kroeber reported for 1859–1864 and that some years had very few cases on which to base measurements. If the data are suspect, Allport and Hartman concluded, then so are the regularities Kroeber claimed to have found (1931:342–43).

Richardson scoured the archives of fashion and extended the database from 1605–1936 (Richardson and Kroeber 1940). Before making measurements for all the new years included in the study, Richardson redid Kroeber's measurements for 1844–1846 and for 1919 and assured herself that she was coding each plate the same way Kroeber had done in 1919 (Richardson and Kroeber 1940).

Lowe and Lowe (1982) reanalyzed the Richardson-Kroeber data for the 150 years from 1787–1936, using all the firepower of modern statistics and computers. You'll be pleased to know that Kroeber's first analysis was vindicated: Stylistic change in women's dress is in stable equilibrium (changing with patterned regularity), and is driven by "inertia, cultural continuity, a rule system of aesthetic proportions, and an inherently unpredictable element" (Lowe and Lowe 1982:521).

Mulcahy and Herbert (1990) added data for the years 1937–1982 and found more variability in those 46 years than in the 150 years before 1937. For example, a plot of the moving average for skirt width from 1811–1926 has the same shape as the plot for 1926–1976. In other words, the cycle of skirt length had been cut by more than half in 1976.

### The Problem with Archival Data

Kroeber's early work is being vindicated, but Allport and Hartman's critique was right on target in 1931. You can't be too critical of your sources. Archival data may appear clean, especially if they come packaged on computer files and are coded and ready to be analyzed. But they may be riddled with error. Consider carefully all the possible sources of bias (informant error, observer error, etc.) that might have been at work in the setting down of the data. Ask how, why, and under what conditions a particular set of archival data was collected. Ask who collected the data and what biases she or he might have had.

No data are free of error. In some parts of Mexico, the number of consensual unions is greater than the number of formal marriages, making court records about marriages problematic. In the United States, on the other hand, crime statistics are notoriously untrustworthy. Many crimes go unreported, and those that are reported may not be recorded at all, or may be recorded in the wrong category. In some countries, rural people may wait as long as 6 months to report a birth, and a significant fraction of their children may die within that period. (See Naroll [1962] and Handlin [1979] for discussions of data quality control in archival research.) It is almost always better to understand distortion in data than to throw them out. (For more on archival research in anthropology, see Brettell 1998.)

## FURTHER READING

*Continuous monitoring*: Algase et al. (1997); Black and Reiss (1967); Chadsey-Rusch and Gonzalez (1988); Drury (1990); Frank et al. (1997); Guilmet (1979); LeCompte (1978); LeCompte et al. (1993); Longabaugh (1980); McCall (1978); Reiss (1971).

*Ethology and human ethology*: Atzwanger et al. (1997); Burkhardt (2005); Chisholm (1983); Eibl-Eiblsfeldt (1989); Houck and Drikamer (1996); Lehner (1996); P. Martin and Bateson (1993).

*Studying children's behavior under natural conditions*: Fine and Sandstrom (1988); McIver et al. (2009); Meehan (2009); Pellegrini (1996).

*Time allocation studies*: Bock (2002); Bove et al. (2000); Gurven and Kaplan (2006); Hames (1992); Umezaki et al. (2002).

*Experience sampling*: Hektner et al. (2007); Scollon et al. (2005). See Kahneman et al. (2004) for a method that duplicates experience sampling through direct interviews rather than with beeper interrupts.

*Pseudopatients and simulated clients*: Chalker et al. (2004); Katz and Naré (2002); Marsh et al. (2004); Tuladhar et al. (1998); Viberg et al. (2009).

*Deception in participant observation*: Brymer (1998); Bulmer (1991); Cassell (1982); Lauder (2003); Lugosi (2006); Rynkiewick and Spradley (1976).

*Proxemics*: Ardener (1981); Høgh-Olesen (2008); Kendon (1981); Kenner and Katsimaglis (1993); Low and Lawrence-Zúñiga (2003).