8

Interviewing I: Unstructured and Semistructured

THE BIG PICTURE

The concept of "interviewing" covers a lot of ground, from totally unstructured interactions, through semistructured situations, to highly formal interactions with respondents. Interviewing is done on the phone, in person, by mail and—more and more—by computer or on the Internet. This chapter is about unstructured and semistructured face-to-face interviewing, including the management of focus groups.

Unstructured interviewing goes on all the time and just about anywhere—in homes, walking along a road, weeding a millet field, hanging out in bars, or waiting for a bus. **Semistructured**, or **in-depth interviewing** is a scheduled activity. A semistructured interview is open ended, but follows a general script and covers a list of topics.

There is a vast literature on how to conduct effective interviews: how to gain rapport, how to get people to open up, how to introduce an interview, and how to end one. You can't learn to interview by reading about it, but after you read this chapter and practice some of the techniques described, you should be well on your way to becoming an effective interviewer. You should also have a pretty good idea of how much more there is to learn, and be on your way to exploring the literature.

INTERVIEW CONTROL

There is a continuum of interview situations based on the **amount of control** we try to exercise over people's responses (Dohrenwend and Richardson 1965; Gorden 1987; Spradley 1979). These *different types* of interviews produce *different types* of data that are useful for *different types* of research projects and that appeal to *different types* of researchers. For convenience, I divide the continuum of interviews into four large chunks.

1. Informal Interviewing

At one end there is **informal interviewing**, characterized by a total lack of structure or control. The researcher just tries to remember conversations heard during the course of a day in the field. This requires constant jotting and daily sessions in which you sit at a computer, typing away, unburdening your memory, and developing field notes. Informal interviewing is the method of choice at the beginning of participant observation fieldwork, when you're settling in. It is also used throughout ethnographic fieldwork to build greater rapport and to uncover new topics of interest that might have been overlooked.

When it comes to interviewing, never mistake the adjective "informal" for "light-

weight." This is hard, hard work. You have to remember a lot; you have to duck into private corners a lot (so you can jot things down); and you have to use a lot of deception (to keep people from knowing that you're really at work, studying them). Informal interviewing can get pretty tiring.

Still, in some kinds of research, informal interviewing is all you've got. Mark Connolly (1990) studied *gamines*, or street children, in Guatemala City, Guatemala, and Bogotá, Colombia. These children live, eat, and sleep on the street. Hanging out and talking informally with these children was an appropriate way to do this research. Informal ethnography can also be combined with more structured methods, when circumstances allow it. In fact, Rachel Baker (1996a, 1996b) was able to collect anthropometric data on street children in Kathmandu, Nepal, while doing informal ethnography.

2. Unstructured Interviewing

Next comes **unstructured interviewing**, one of the two types covered in this chapter. There is nothing at all informal about unstructured interviewing, and nothing deceptive, either. You sit down with another person and hold an interview. Period. Both of you know what you're doing, and there is no shared feeling that you're just engaged in pleasant chitchat.

Unstructured interviews are based on a clear plan that you keep constantly in mind, but are also characterized by a minimum of control over the people's responses. The idea is to get people to open up and let them express themselves in their own terms, and at their own pace. A lot of what is called **ethnographic interviewing** is unstructured. Unstructured interviewing is used in situations where you have lots and lots of time—like when you are doing long-term fieldwork and can interview people on many separate occasions (box 8.1).

BOX 8.1

PAYING INFORMANTS

Should anthropologists pay their informants? If so, how much? I'm a firm believer in paying for people's time, but there are exceptions. If you are studying people who are worth millions of dollars, paying them is inappropriate. You can't possibly pay them enough to compensate them financially for their time. It's better to make a donation to a charity that they support. This will vary from case to case, but the general rule, for me at least, is that if you want to interview people formally—sit down with them, voice recorder on the table and/or notebook in hand—they should be paid at least the local rate for their time. With key informants, the rule for me is that there's always a culturally appropriate way—money, job training, buying cement for a new school—to compensate people for their contribution to your career.

3. Semistructured Interviewing

In situations where you won't get more than one chance to interview someone, **semistructured interviewing** is best. It has much of the freewheeling quality of unstructured

interviewing and requires all the same skills, but semistructured interviewing is based on the use of an **interview guide**. This is a written list of questions and topics that need to be covered in a particular order.

This is the kind of interview that most people write about—the kind done in professional surveys. The interviewer maintains discretion to follow leads, but the interview guide is a set of clear instructions—instructions like this one: "Probe to see if informants (men and women alike) who have daughters have different values about dowry and about premarital sex than do people who have only sons."

Formal, written guides are an absolute must if you are sending out several interviewers to collect data. But even if you do all the interviewing on a project yourself, you should build a guide and follow it if you want reliable, comparable qualitative data.

Semistructured interviewing works very well in projects where you are dealing with high-level bureaucrats and elite members of a community—people who are accustomed to efficient use of their time. It demonstrates that you are fully in control of what you want from an interview but leaves both you and your respondent to follow new leads. It shows that you are prepared and competent but that you are not trying to exercise excessive control.

4. Structured Interviewing

Finally, in fully **structured interviews**, people are asked to respond to as nearly identical a set of stimuli as possible. One variety of structured interviews involves use of an **interview schedule**—an explicit set of instructions to interviewers who administer questionnaires orally. Instructions might read: "If the informant says that she or he has at least one daughter over 10 years of age, then ask questions 26b and 26c. Otherwise, go on to question 27."

Self-administered questionnaires are a kind of structured interview. Other structured interviewing techniques include pile sorting, frame elicitation, triad sorting, and tasks that require informants to rate or rank order a list of things. I'll deal with these in chapter 10.

UNSTRUCTURED INTERVIEWING

Unstructured interviewing is truly versatile. It is used equally by scholars who identify with the hermeneutic tradition and by those who identify with the positivist tradition. It is used in studies that require only textual data and in studies that require both textual and numerical data. Ethnographers may use it to develop formal guides for semistructured interviews, or to learn what questions to include, in the native language, on a highly structured questionnaire (see Werner and Schoepfle [1987] for a good discussion of this). I say that ethnographers *may* use unstructured interviewing in developing structured interview schedules because unstructured interviewing also stands on its own.

When you want to know about the **lived experience** of fellow human beings—what it's like to survive hand-to-hand combat, how you get through each day when you have a child dying of leukemia, how it feels to make it across the border into Texas from Mexico only to be deported 24 hours later—you just can't beat unstructured interviewing.

Unstructured interviewing is excellent for building initial rapport with people, before moving to more formal interviews, and it's perfect for talking to informants who would not tolerate a more formal interview. The personal rapport you build with close informants in long-term fieldwork can make highly structured interviewing—and even semistructured interviewing—feel somehow unnatural. In fact, really structured interviewing can get in the way of your ability to communicate freely with key informants.

But not always. Some people want very much to talk about their lives, but they really don't like the unstructured interview format. I once asked a fisherman in Greece if I could have a few minutes of his time to discuss the economics of small-scale fishing. I was about 5 minutes into the interview, treading lightly—you know, trying not to get too quickly into his finances, even though that's exactly what I wanted to know about—when he interrupted me: "Why don't you just get to the point?" he asked. "You want to know how I decide where to fish, and whether I use a share system or a wage system to split the profits, and how I find buyers for my catch, and things like that, right?" He had heard from other fishermen that these were some of the topics I was interviewing people about. No unstructured interviews for him; he was a busy man and wanted to get right to it.

A Case Study of Unstructured Interviewing

Once you learn the art of probing (which I'll discuss in a bit), unstructured interviewing can be used for studying sensitive issues, like sexuality, racial or ethnic prejudice, or hot political topics. I find it particularly useful in studying conflict. In 1972–1973, I went to sea on two different oceanographic research vessels (Bernard and Killworth 1973, 1974). In both cases, there was an almost palpable tension between the scientific personnel and the crew of the ship. Through both informal and unstructured interviewing on land between cruises, I was able to establish that the conflict was predictable and regular. Let me give you an idea of how complex the situation was.

In 1972–1973, it cost $5,000 a day to run a major research vessel, not including the cost of the science. (That would be about $26,000 today.) The way oceanography works, at least in the United States, the chief scientist on a research cruise has to pay for both ship time and for the cost of any experiments he or she wants to run. To do this, oceanographers compete for grants from institutions like the U.S. Office of Naval Research, NASA, and the National Science Foundation.

The spending of so much money is validated by publishing significant results in prominent journals. It's a tough, competitive game and one that leads scientists to use every minute of their ship time. As one set of scientists comes ashore after a month at sea, the next set is on the dock waiting to set up their experiments and haul anchor.

The crew, consequently, might only get 24 or 48 hours shore leave between voyages. That can cause some pretty serious resentment by ships' crews against scientists. And that can lead to disaster. I found many documented instances of sabotage of expensive research by crew members who were, as one of them said, "sick and tired of being treated like goddamn bus drivers." In one incident, involving a British research vessel, a freezer filled with Antarctic shrimp, representing 2 years of data collection, went overboard during the night. In another, the crew and scientists from a U.S. Navy oceanographic research ship got into a brawl while in port (*Science* 1972:1346).

The structural problem I uncovered began at the top. Scientists whom I interviewed felt they had the right to take the vessels wherever they wanted to go, within prudence and reason, in search of answers to questions they had set up in their proposals. The captains of the ships believed (correctly) that *they* had the last word on maneuvering their ships at sea. Scientists, said the captains, sometimes went beyond prudence and reason in what they demanded of the vessels.

For example, a scientist might ask the captain to take a ship out of port in dangerous weather because ship time is so precious. This conflict between crew and scientists has been known—and pretty much ignored—since Charles Darwin sailed with HMS *Beagle* and it will certainly play a role in the productivity of long-term space station operations.

Unraveling this conflict at sea required participant observation and unstructured (as

well as informal) interviewing with many people. No other strategy for data collection would have worked. At sea, people live for weeks, or even months, in close quarters, and there is a common need to maintain good relations for the organization to function well.

It would have been inappropriate for me to have used highly structured interviews about the source of tension between the crew and the scientists. Better to steer the interviews around to the issue of interest and to let informants teach me what I needed to know. In the end, no analysis was better than that offered by one engine room mechanic who told me, "These scientist types are so damn hungry for data, they'd run the ship aground looking for interesting rocks if we let them."

Getting Started

There are some important steps to take when you start interviewing someone for the first time. First of all, assure people of anonymity and confidentiality. Explain that you simply want to know what *they* think, and what *their* observations are. If you are interviewing someone whom you have come to know over a period of time, explain why you think their opinions and observations on a particular topic are important. If you are interviewing someone chosen from a random sample, and whom you are unlikely to see again, explain how they were chosen and why it is important that you have their cooperation to maintain representativeness.

If people say that they really don't know enough to be part of your study, assure them that their participation is crucial and that you are truly interested in what they have to say (and you'd better mean it or you'll never pull it off). Tell everyone you interview that you are trying to learn from *them*. Encourage them to interrupt you during the interview with anything they think is important. And always ask for permission to *record* personal interviews *and to take notes*. This is vital. If you can't take notes, then, in most cases, the value of an interview plummets. (See below, on using a recorder and taking notes.)

Keep in mind that people who are being interviewed know that you are shopping for information. There is no point in trying to hide this. If you are open and honest about your intentions, and if you are genuinely interested in what people have to say, many people will help you.

This is not always true, though. When Colin Turnbull went out to study the Ik in Uganda, he found a group of people who had apparently lost interest in life and in exchanging human kindnesses. The Ik had been brutalized, decimated, and left by the government to fend for themselves on a barren reservation. They weren't impressed with the fact that Turnbull wanted to study their culture. In fact, they weren't much interested in anything Turnbull was up to and were anything but friendly (Turnbull 1972).

Letting the Informant or Respondent Lead

If you can carry on "unthreatening, self-controlled, supportive, polite, and cordial interaction in everyday life," then interviewing will come easy to you, and informants will feel comfortable responding to your questions (Lofland 1976:90). But no matter how supportive you are as a person, an interview is never really like a casual, unthreatening conversation in everyday life. In casual conversations, people take more or less balanced turns (Spradley 1979) and there is no feeling that somehow the discussion has to stay on track or follow some theme (see also Hyman and Cobb 1975; Merton et al. 1956). In unstructured interviewing, you keep the conversation focused on a topic, while giving the respondent room to define the content of the discussion.

The rule is: Get people on to a topic of interest and get out of the way. Let the informant provide information that he or she thinks is important.

During my research on the Kalymnian sponge fishermen in Greece, I spent a lot of time at Procopis Kambouris's *taverna.* (A Greek *taverna* is a particular kind of restaurant.) Procopis's was a favorite of the sponge fishermen. Procopis was a superb cook, he made his own wine every year from grapes that he selected himself, and he was as good a teller of sea stories as he was a listener to those of his clientele. At Procopis's *taverna*, I was able to collect the work histories of sponge fishermen—when they'd begun their careers, the training they'd gotten, the jobs they'd held, and so on. The atmosphere was relaxed (plenty of retsina wine and good things to eat), and conversation was easy.

As a participant observer, I developed a sense of camaraderie with the regulars, and we exchanged sea stories with a lot of flourish. Still, no one at Procopis's ever made the mistake of thinking that I was there just for the camaraderie. They knew that I was writing about their lives and that I had lots of questions to ask. They also knew immediately when I switched from the role of participant observer to that of ethnographic interviewer.

One night, I slipped into just such an interview/conversation with Savas Ergas. He was 64 years old at the time and was planning to make one last 6-month voyage as a sponge diver during the coming season in 1965. I began to interview Savas on his work history at about 7:30 in the evening, and we closed Procopis's place at about 3 in the morning. During the course of the evening, several other men joined and left the group at various times, as they would on any night of conversation at Procopis's. Savas had lots of stories to tell (he was a living legend and he played well to a crowd), and we had to continue the interview a few days later, over several more liters of retsina.

At one point on that second night, Savas told me (almost offhandedly) that he had spent more than a year of his life walking the bottom of the Mediterranean. I asked him how he knew this, and he challenged me to document it. Savas had decided that there was something important that I needed to know and he maneuvered the interview around to make sure I learned it.

This led to about 3 hours of painstaking work. We counted the number of seasons he'd been to sea over a 46-year career (he remembered that he hadn't worked at all during 1943 because of "something to do with the war"). We figured conservatively the number of days he'd spent at sea, the average number of dives per trip, and the average depth and time per dive. We joked about the tendency of divers to exaggerate their exploits and about how fragile human memory is when it comes to this kind of detail.

It was difficult to stay on the subject, because Savas was such a good raconteur and a perceptive analyst of Kalymnian life. The interview meandered off on interesting tangents, but after a while, either Savas or I would steer it back to the issue at hand. In the end, discounting heavily for both exaggeration and faulty recall, we reckoned that he'd spent at least 10,000 hours—about a year and a fourth, counting each day as a full 24 hours—under water and had walked the distance between Alexandria and Tunis at least three times.

The exact numbers really didn't matter. What did matter was that Savas Ergas had a really good sense of what *he* thought I needed to know about the life of a sponge diver. It was I, the interviewer, who defined the focus of the interview; but it was Savas, the respondent, who determined the content. And was I ever glad he did.

PROBING

The key to successful interviewing is learning how to probe effectively—that is, to stimulate a respondent to produce more information, without injecting yourself so much into the interaction that you only get a reflection of yourself in the data. Suppose you ask, "Have you ever been away from the village to work?" and the informant says, "Yes." The

next question (the probe) is: "Like where?" Suppose the answer is, "Oh, several different places." The correct response is not, "Pachuca? Querétaro? Mexico City?" but, "Like where? Could you name some of the places where you've gone to get work?"

There are many kinds of probes that you can use in an interview. (In what follows, I draw on the important work by Briggs [1986], Dohrenwend and Richardson [1965], Gorden [1987], Hyman and Cobb [1975], Kahn and Cannell [1957], Kluckhohn [1945], Merton et al. [1956], Reed and Stimson [1985], Warwick and Lininger [1975], Whyte [1960], Whyte and Whyte [1984], and on my own experience and that of my students.)

The Silent Probe

The most difficult technique to learn is the **silent probe**, which consists of just remaining quiet and waiting for an informant to continue. The silence may be accompanied by a nod or by a mumbled "uh-huh" as you focus on your note pad. The silent probe sometimes produces more information than does direct questioning. At least at the beginning of an interview, informants look to you for guidance as to whether or not they're on the right track. They want to know whether they're "giving you what you want." Most of the time, especially in unstructured interviews, you want the informant to define the relevant information.

Some informants are more glib than others and require very little prodding to keep up the flow of information. Others are more reflective and take their time. Inexperienced interviewers tend to jump in with verbal probes as soon as an informant goes silent. Meanwhile, the informant may be just reflecting, gathering thoughts, and preparing to say something important. You can kill those moments (and there are a lot of them) with your interruptions.

Glibness can be a matter of *cultural*, not just personal style. Gordon Streib reports that he had to adjust his own interviewing style radically when he left New York City to study the Navajo in the 1950s (Streib 1952). Streib, a New Yorker himself, had done studies based on semistructured interviews with subway workers in New York. Those workers maintained a fast, hard-driving pace during the interviews—a pace with which Streib, as a member of the culture, was comfortable.

But that style was entirely inappropriate with the Navajo, who were uniformly more reflective than the subway workers (Streib, personal communication). In other words, the silent probe is sometimes not a "probe" at all; being quiet and waiting for an informant to continue may simply be appropriate cultural behavior.

On the other hand, the silent probe is a high-risk technique, which is why beginners avoid it. If an informant is genuinely at the end of a thought and you don't provide further guidance, your silence can become awkward. You may even lose your credibility as an interviewer. The silent probe takes practice to use effectively. But it's worth the effort.

The Echo Probe

Another kind of probe consists of simply repeating the last thing someone has said, and asking them to continue. This **echo probe** is particularly useful when an informant is describing a process, or an event. "I see. The goat's throat is cut and the blood is drained into a pan for cooking with the meat. Then what happens?" This probe is neutral and doesn't redirect the interview. It shows that you understand what's been said so far and encourages the informant to continue with the narrative. If you use the echo probe too often, though, you'll hear an exasperated informant asking, "Why do you keep repeating what I just said?"

The Uh-huh Probe

You can encourage an informant to continue with a narrative by just making affirmative comments, like "Uh-huh," or "Yes, I see," or "Right, uh-huh," and so on. Matarazzo (1964) showed how powerful this **neutral probe** can be. He did a series of identical, semistructured, 45-minute interviews with a group of informants. He broke each interview into three 15-minute chunks. During the second chunk, the interviewer was told to make affirmative noises, like "uh-huh," whenever the informant was speaking. Informant responses during those chunks were about a third longer than during the first and third periods. If you use the uh-huh or the silent probe, be sure to stay engaged and look directly at your informants.

The Tell-Me-More Probe

This may be the most common form of probe among experienced interviewers. Respondents give you an answer, and you probe for more by saying: "Could you tell me more about that?" Other variations include "Why exactly do you say that?" and "Why exactly do you feel that way?" You have to be careful about using stock probes like these. As Converse and Schuman point out (1974:50), if you get into a rut and repeat these probes like a robot, don't be surprised to hear someone finishing up a nice long discourse by saying, "Yeah, yeah, and why *exactly* do I feel like that?" (I can guarantee personally that the mortification factor only allows this sort of thing to happen once. The memory of the experience lasts a lifetime.)

The Long Question Probe

Another way to induce longer and more continuous responses is by making your questions longer. Instead of asking, "How do you plant a home garden?" ask, "What are all the things you have to do to actually get a home garden going?" When I interviewed sponge divers on Kalymnos, instead of asking them, "What is it like to make a dive into very deep water?" I said, "Tell me about diving into really deep water. What do you do to get ready and how do you descend and ascend? What's it like down there?"

Later in the interview or on another occasion, I would home in on special topics. But to break the ice and get the interview flowing, there is nothing quite as useful as what Spradley (1979) called the **grand tour question.**

This does not mean that asking longer questions or using neutral probes necessarily produces *better* responses. They do, however, produce *more* responses, and, in general, more is better. Furthermore, the more you can keep an informant talking, the more you can express interest in what they are saying and the more you build rapport. This is especially important in the first interview you do with someone whose trust you want to build (see Spradley 1979:80). There is still a lot to be learned about how various kinds of probes affect what informants tell us.

Threatening questions—those asking for sensitive information—should be short but preceded by a long, rambling run-up: "We're interested in the various things that people do these days in order to keep from getting diseases when they have sex. Some people do different kinds of things, and some people do nothing special. Do you ever use condoms?" If the respondents says, "Yes," or "No," or "Sometimes," *then* you can launch that series of questions about why, why not, when, with whom, and so on. The wording of sensitive questions should be supportive and nonjudgmental. (See below for more on threatening questions.)

Probing by Leading

After all this, you may be cautious about being really directive in an interview. Don't be. Many researchers caution against "leading" an informant. Lofland (1976), for example, warns against questions like, "Don't you think that . . . ?" and suggests asking, "What do you think about . . . ?" He is, of course, correct. On the other hand, any question an interviewer asks leads an informant. You might as well learn to do it well.

Consider this leading question that I asked a Ñähñu Indian: "Right. I understand. The compadre is *supposed* to pay for the music for the baptism fiesta. But what happens if the compadre doesn't have the money? Who pays then?" This kind of question can stop the flow of an informant's narrative stone dead. It can also produce more information than the informant would otherwise have provided. At the time, I thought the informant was being overly "normative." That is, I thought he was stating an ideal behavioral custom (having a compadre pay for the music at a fiesta) as if it were never violated.

It turned out that all he was doing was relying on his own cultural competence—"abbreviating," as Spradley (1979:79) called it. The informant took for granted that the anthropologist knew the "obvious" answer: If the compadre didn't have enough money, well, then there might not be any music.

My interruption reminded the informant that I just wasn't up to his level of cultural competence; I needed him to be more explicit. He went on to explain other things that he considered obvious but that I would not have even known to ask about. Someone who has committed himself to pay for the music at a fiesta might borrow money from *another* compadre to fulfill the obligation. In that case, he wouldn't tell the person who was throwing the fiesta. That might make the host feel bad, like he was forcing his compadre to go into debt.

In this interview, in fact, the informant eventually became irritated with me because I asked so many things that he considered obvious. He wanted to abbreviate a lot and to provide a more general summary: I wanted details. I backed off and asked a different informant for the details. I have since learned to start some probes with "This may seem obvious, but . . ." (box 8.2).

BOX 8.2

LISTEN FOR WHAT'S LEFT OUT

Informants abbreviate all the time, and this means that you have to listen carefully for what's *left out,* not just what's *in* the interview narrative. Laurie Price (1987) collected tales of misfortune from very poor people in Quito, Ecuador. In one story, Maria talks about her crippled 6-year-old daughter. As Price tells it, Maria does not mention that, for months, she carried her daughter every day "down a 200-step flight of public stairs and 4 blocks to the nearest bus stop so the girl could go to physical therapy" (p. 318). The child's father, it turns out, drives a bus that he parks every night next to their house, but during the Herculean effort to help the daughter, the father never pitches in or rearranges his schedule. "Such efforts are the unmarked case for mothers," says Price (p. 319).

Directive probes (leading questions) may be based on what an informant has just finished saying, or may be based on something an informant told you an hour ago or a

week ago. As you progress in long-term research, you come to have a much greater appreciation for what you really want from an interview. It is perfectly legitimate to use the information you've already collected to focus your subsequent interviews.

This leads researchers from informal to unstructured to semistructured interviews and even to completely structured interviews like questionnaires. When you feel as though you have learned something important about a group and its culture, the next step to test that knowledge—to see if it is idiosyncratic to a particular informant or subgroup in the culture or if it can be reproduced in many informants.

Baiting: The Phased-Assertion Probe

A particularly effective probing technique is called **phased assertion** (Kirk and Miller 1986), or **baiting** (Agar 1996:142). This is when you act like you already know something to get people to open up.

I used this technique in a study of how Ñähñu Indian parents felt about their children learning to read and write Ñähñu. Bilingual (Spanish-Indian) education in Mexico is a politically sensitive issue (Heath 1972), and when I started asking about it a lot of people were reluctant to talk freely.

In the course of informal interviewing, I learned from a schoolteacher in one village that some fathers had come to complain about the teacher trying to get the children to read and write Ñähñu. The fathers, it seems, were afraid that studying Ñähñu would get in the way of their children becoming fluent in Spanish. Once I heard this story, I began to drop hints that I knew the reason parents were against children learning to read and write Ñähñu. As I did this, the parents opened up and confirmed what I'd found out.

Every journalist (and gossip monger) knows this technique well. As you learn a piece of a puzzle from one informant, you use it with the next informant to get more information, and so on. The more you seem to know, the more comfortable people feel about talking to you and the less people feel they are actually divulging anything. *They* are not the ones who are giving away the "secrets" of the group.

Phased assertion also prompts some informants to jump in and correct you if they think you know a little but that you've "got it all wrong." In some cases, I've purposely made wrong assertions to provoke a correcting response.

Verbal Respondents

Some people try to tell you *too much*. They are the kind of people who just love to have an audience. You ask them one little question and off they go on one tangent after another, until you become exasperated. Converse and Schuman (1974:46) recommend "gentle inattention"—putting down your pen, looking away, leafing through your papers. Nigel King (1994:23) recommends saying something like: "That's very interesting. Could we go back to what you were saying earlier about . . ."

You may, however, have to be a bit more obvious. New interviewers, in particular, may be reluctant to cut off informants, afraid that doing so is poor interviewing technique. In fact, as William Foote Whyte notes, informants who want to talk your ear off are probably used to being interrupted. It's the only way their friends get a word in edgewise. But you need to learn how to cut people off without rancor. "Don't interrupt *accidentally* . . . ," Whyte said, "learn to interrupt *gracefully*" (1960:353, emphasis his). Each situation is somewhat different; you learn as you go in this business.

Nonverbal Respondents

One of the really tough things you run into is someone telling you "I don't know" in answer to lots of questions. In qualitative research projects, where you choose respondents

precisely because you think they know something of interest, the "don't know" refrain can be especially frustrating. Converse and Schuman (1974:49) distinguish four kinds of don't-know response: (1) I don't know (and frankly I don't care); (2) I don't know (and it's none of your business); (3) I don't know (actually, I do know, but you wouldn't be interested in what I have to say about that); and (4) I don't know (and I wish you'd change the subject because this line of questioning makes me really uncomfortable). There is also the "(I wish I could help you but) I really don't know."

Sometimes you can get beyond this, sometimes you can't. You have to face the fact that not everyone who volunteers to be interviewed is a good respondent. If you probe those people for information when they say, "I don't know," you tempt them to make something up just to satisfy you, as Sanchez and Morchio (1992) found. Sometimes, you just have to take the "don't know" for an answer and cut your losses by going on to someone else (box 8.3).

LEARNING TO INTERVIEW

It's impossible to eliminate reactivity and subjectivity in interviewing, but like any other craft, you get better and better at interviewing the more you practice. It helps a lot to practice in front of others and to have an experienced interviewer monitor and criticize your performance. Even without such help, however, you can improve your interviewing technique just by paying careful attention to what you're doing. Harry Wolcott (1995) offers excellent advice on this score: Pay as much attention to your own words as you do to the words of your respondents (p. 102).

Wolcott also advises: Keep interviews focused on a few big issues (1995:112). More good advice from one of the most accomplished ethnographers around. Here's a guaranteed way to wreck rapport and ruin an interview: An informant asks you, "Why do you ask? What does that have to do with what we're talking about?" You tell her: "Well, it just seemed like an interesting question—you know, something I thought might be useful somehow down the road in the analysis."

Here you are, asking people to give you their time and tell you about their lives and you're treating that time with little respect. If you can't imagine giving a satisfactory answer to the question: "Why did you ask *that*?" then leave *that* out.

Do *not* use your friends as practice informants. You cannot learn to interview with friends because there are role expectations that get in the way. Just when you're really rolling, and getting into probing deeply on some topic that you both know about, they are likely to laugh at you or tell you to knock it off.

Practice interviews should *not* be just for practice. They should be done on topics you're really interested in and with people who are likely to know a lot about those topics. Every interview you do should be conducted as professionally as possible and should produce useful data (with plenty of notes that you can code, file, and cross-file).

The Importance of Language

Most anthropologists (and an increasing number of sociologists and social psychologists) do research outside their own country. If you are planning to go abroad for research, find people from the culture you are going to study and interview them on some topic of interest. If you are going to Turkey to study women's roles, then find Turkish students at your university and interview them on some related topic.

It is often possible to hire spouses of foreign students for these kinds of "practice" interviews. I put "practice" in quotes to emphasize again that these interviews should

BOX 8.3

THE ETHICS OF PROBING

Are these tricks of the trade ethical? Peter Collings (2009) asked Inuit hunters: "Name all of the people you share country food with." The response was usually a very short list, so, when informants stopped listing names, Collings would ask "What about your x? Surely you share food with your x," where x was a category of relative in Innuinaqtun, the local language. This, said Collings, reminded people that he had command of the kinship terminology and that he knew his informant had an x. Later in the interview, Collings would refer to one of the people whom the informant had named and say "So-and-so is your older brother," using the Innuinaqtun term. "Just as the fieldworker is studying the community," says Collings, "so, too, is the community studying the fieldworker" to find out if he or she is culturally competent (pp. 149–50). By demonstrating cultural competence, Collings argues, phased assertion helps establish rapport—at least where he works (p. 139).

Still, getting people to open up creates responsibilities for your informants. First, there is no ethical imperative in social research more important than seeing to it that you do not harm innocent people who have provided you with information in good faith. Not all respondents are innocents, though. Some people commit wartime atrocities. Some practice infanticide. Some are HIV-positive and, out of bitterness, are purposely infecting others. Do you protect them all? Are any of these examples more troublesome to you than others? These are not extreme cases, thrown in here to prepare you for the worst, "just in case." They are the sorts of ethical dilemmas that field researchers confront all the time.

Second, the better you get at making people open up, the more responsible you become that they don't later suffer some emotional distress for having done so. Informants who divulge *too* quickly what they believe to be secret information can later come to have real regrets and even loss of self-esteem. They may suffer anxiety over how much they can trust you to protect them in the community.

It is sometimes better to stop an informant from divulging privileged information in the first or second interview and to wait until both of you have built a mutually trusting relationship. If you sense that an informant is uncomfortable with having spoken too quickly about a sensitive topic, end the interview with light conversation and reassurances about your discretion. Soon after, look up the informant and engage in light conversation again, with no probing or other interviewing techniques involved. This will also provide reassurance of trust.

Remember: The first ethical decision you make in research is whether to collect certain kinds of information at all. Once that decision is made, *you* are responsible for what is done with that information, and *you* must protect people from becoming emotionally burdened for having talked to you.

produce real data of real interest to you. If you are studying a language that you'll need for research, these practice interviews will help you sharpen your skills at interviewing in that language.

Even if you are going off to the interior of the Amazon, this doesn't let you off the hook. It is unlikely that you'll find native speakers of Yanomami on your campus, but you cannot use this as an excuse to wait until you're out in the field to learn general interviewing skills. Interviewing skills are honed by practice. Among the most constructive things you can do in preparing for field research is to practice conducting unstructured and semistructured interviewing. Learn to interview in Portuguese or Spanish (depending on whether the Yanomami you are going to visit live in the Brazilian or Venezuelan Amazon) before heading for the field and you'll be way ahead. (See the section on language in chapter 12 for more on using interpreters.)

Pacing the Study

Two of the biggest problems faced by researchers who rely heavily on semistructured interviews are boredom and fatigue. Even small projects may require 30–40 interviews to generate sufficient data to be worthwhile. Most field researchers collect their own interview data, and asking the same questions over and over again can get pretty old. Gorden (1987) studied 30 interviewers who worked for 12 days doing about two tape-recorded interviews per day. Each interview was from 1 to 2 hours long.

The first interview on each day, over all interviewers, averaged about 30 pages of transcription. The second averaged only 25 pages. Furthermore, the first interviews, on average, got shorter and shorter during the 12-day period of the study. In other words, on any given day, boredom made the second interview shorter, and over the 12 days boredom (and possibly fatigue) took its toll on the first interviews of each day.

Even anthropologists who spend a year in the field may have focused bouts of interviewing on a particular topic. Plan each project, or subproject, in advance and calculate the number of interviews you are going to get. Pace yourself. Spread the project out if possible, and don't try to bring in all your interview data in the shortest possible time—unless you're studying reactions to a hot issue, in which case, spreading things out can create a serious history confound (see chapter 4).

Here's the tradeoff: The longer a project takes, the less likely that the first interviews and the last interviews will be valid indicators of the same things. In long-term, participant observation fieldwork (6 months to a year), I recommend going back to your early informants and interviewing them a second time. See whether their observations and attitudes have changed, and if so, why.

PRESENTATION OF SELF

How should you present yourself in an interview? As a friend? As a professional? As someone who is sympathetic or as someone who is nonjudgmental? It depends on the nature of the project. When the object is to collect comparable data across respondents, then it makes no difference whether you're collecting words or numbers—cordial-but-nonjudgmental is the way to go.

That's sometimes tough to do. You're interviewing someone on a project about what people can do to help the environment, and your respondent says: "All those eco-Nazis want is to make room for more owls. They don't give a damn about real people's jobs." (Yes, that happened on one of my projects.) That's when you find out whether you can probe without injecting your feelings into the interview. Professional interviewers (the

folks who collect the data for the General Social Survey, for example) learn to maintain their equilibrium and move on (see Converse and Schuman 1974).

Some situations are so painful, however, that it's impossible to maintain a neutral facade. Gene Shelley interviewed 72 people in Atlanta, Georgia, who were HIV-positive (Shelley et al. 1995). Here's a typical comment by one of Shelly's informants: "I have a lot of trouble watching all my friends die. Sometimes my whole body shuts down inside. I don't want to know people who are going to die. Some of my friends, there are three or four people a week in the obits. We all watch the obits."

How would *you* respond? Do you say: "Uh-huh. Tell me more about that"? Do you let silence take over and force the respondent to go on? Do you say something sympathetic? Shelley reports that she treated each interview as a unique situation and responded as her intuition told her to respond—sometimes more clinically, sometimes less, depending on her judgment of what the respondent needed her to say. Good advice.

On Just Being Yourself

In 1964, when we were working on the island of Kalymnos, my wife Carole would take our 2-month-old baby for daily walks in a carriage. Older women would peek into the baby carriage and make disapproving noises when they saw our daughter sleeping on her stomach. Then they would reach into the carriage and turn the baby over, explaining forcefully that the baby would get the evil eye if we continued to let her sleep on her stomach.

Carole had read the latest edition of *The Commonsense Book of Baby and Child Care* (the classic "baby book" by Dr. Benjamin Spock). We carried two copies of the book with us—in case one fell out of a boat or something—and Carole was convinced by Dr. Spock's writings that babies who sleep on their backs risk choking on their own mucous or vomit. Since then, medical opinion—and all the baby books that young parents read nowadays—have flip-flopped about this issue several times. At the time, though, not wanting to offend anyone, Carole listened politely and tried to act nonjudgmental.

One day, enough was enough. Carole told off a woman who intervened and that was that. From then on, women were more eager to discuss child-rearing practices in general. When we let our baby crawl around on the floor and didn't bundle her up when we took her out for walks, Greek mothers were unhesitant in telling us that they disapproved. The more we challenged them, the more they challenged us. There was no rancor involved, and we learned a lot more than if Carole had just kept on listening politely and had said nothing. This was informal interviewing in the context of long-term participant observation. If we had offended anyone, there would have been time and opportunity to make amends—or at least come to an understanding about cultural differences.

Little Things Mean a Lot

Little things are important in interviewing, so pay attention to them. How you dress and where you hold an interview, for example, tell your respondent a lot about you and what you expect. The "interviewing dress code" is: Use common sense. Proper dress depends on the venue. Showing up with a backpack or an attaché case, wearing jeans or a business suit—these are choices that should be pretty easy to make once you've made the commitment to accommodate your dress to different circumstances.

Same goes for venue. I've held interviews in bars, in business offices, in government offices, on ferry boats, on beaches, in homes. . . . I can't give you a rule for selecting the single *right* place for an interview, since there may be several right places. But some places

are just plain wrong for certain interviews. Here again, common sense goes a long way (**Further Reading:** interviewing).

USING A VOICE RECORDER

Don't rely on your memory in interviewing; use a voice recorder in all structured and semistructured interviews, except where people specifically ask you not to. If you sense some reluctance about the recorder, leave it on the table and don't turn it on right away. Start the interview with chitchat and when things get warmed up, say something like "This is really interesting. I don't want to trust my memory on something as important as this; do you mind if I record it?" Charles Kadushin (personal communication) hands people a microphone with a shut-off switch. Rarely, he says, do respondents actually use the switch, but giving people control over the interview shows that you take them very seriously.

Sometimes you'll be recording an interview and things will be going along just fine and you'll sense that a respondent is backing off from some sensitive topic. Just reach over to the recorder and ask the respondent if she or he would like you to turn it off. Harry Wolcott (1995:114) recommends leaving the recorder on, if possible, when the formal part of an interview ends. Even though you've finished, Wolcott points out, your respondent may have more to say.

Recording Equipment

For simple recording and transcribing of interviews, in a language you understand well, you can get away with a basic audio recorder for under $50. (But buy two of them. When you skimp on equipment costs, and don't have a spare, this almost guarantees that you'll need one at the most inconvenient moment.) Basic recorders, with 256mb of flash memory hold about 150 hours of voice recording. You can also use your iPod as a digital audio recorder with a plug-in microphone. A gigabyte of disk space holds about 400 hours of voice recordings, so an 80-gigabyte iPod has plenty of room for both music and interviews.

Whatever kind of work you do, remember to upload your data regularly to a computer and to store your data in several places—CDs, external hard drives, or online. And if you are in an isolated field site and don't have reliable power, take along a solar battery charger so you can get your data offline and onto a CD.

Some of the better voice recorders come with up to four built-in microphones that capture 360-degree sound. If you use a low-end recorder, then use a good, separate microphone. Some people like wearing a lavalier microphone—the kind you clip to a person's lapel or shirt collar—but many people find them intrusive. I prefer omnidirectional microphones because they pick up voices from anywhere in a room. Sometimes, people get rolling on a topic and they want to get up and pace the room as they talk. Want to kill a really great interview? Tell somebody who's on a roll to please sit down and speak directly into the mike. Good microphones come with stands that keep the head from resting on any surface, like a table. Surfaces pick up and introduce background noise into any recording. If you don't have a really good stand for the mike, you can make one easily with some rubbery foam (the kind they use in making mattresses).

Test your recorder before every interview. And do the testing at home. There's only one thing worse than a recorder that doesn't run at all. It's one that runs but doesn't record. Then your informant is sure to say at the end of the interview: "Let's run that back and see how it came out!" (Yes, that happened to me. But only once. And it needn't happen to anyone who reads this.)

Pay attention to the battery indicator. Want another foolproof way to kill an exciting interview? Ask the informant to "please hold that thought" while you change batteries. When batteries get slightly low, throw them out or recharge them. If you are working in places that have unstable current, you'll rely on batteries to ensure recording fidelity. Just make sure that you start out with fresh batteries for each interview. (You can save a lot of battery life by using house current for all playback, fast forward, and rewind operations—reserving the batteries *only* for recording.) If you prefer household current for recording, then carry along a couple of long extension cords so you have a choice of where to set up for the interview. (See Ives [1995] for more good tips.)

In voice activation mode, the recorder only turns off during long pauses—while an informant is thinking, for example. Holly Williams (personal communication) recommends not using the voice activation feature for interviews. She finds that the long breaks without any sound make transcribing easier. You don't have to shut the machine off and turn it on as many times while you're typing.

Transcribers and VR software

It takes 6–8 hours to transcribe 1 hour of a recorded interview, depending on how closely you transcribe (getting all the "uhs" and "ers" and throat clearings, or just capturing the main elements of speech), how clear the recording is, and how proficient you are in the language and in typing. If you have to transcribe interviews yourself, there are several choices for equipment. Transcription software lets you control the recorder (start, stop, move forward and backward) using the keyboard. Transcriber machines let you do this using a foot pedal. This lets you listen to a couple of seconds of recording at a time, type everything into the computer, and then move on to the next chunk. The technology lets you go back and repeat chunks, all while keeping your hands on the keyboard.

With voice recognition (VR) software, you listen to an interview through a set of headphones and repeat the words—both your questions and your informant's responses—out loud, in your own voice. The software listens to your voice and types out the words across the screen. You go over each sentence to correct mistakes (tell it that the word "bloat" should be "float" for instance) and to format the text (tell it where to put punctuation and paragraph breaks). The process is slow at first, but the software learns over time to recognize inflections in your voice, and it makes fewer and fewer mistakes as weeks go by. It also learns all the special vocabulary you throw at it. The built-in vocabularies of current VRS systems are enormous—something like 300,000 words—but, though they may be ready to recognize polygamy, for example, you'll have to teach it polygyny or fraternal polyandry and words from the language of your field site. So, if you say, "Juanita sold eight *huipiles* at the market this week," you'll have to spell out "Juanita" and "*huipiles*" so the software can add these words to its vocabulary.

As the software gets trained, the process moves up to 95%–98% accuracy at about 100–120 words per minute. With a 2%–5% error rate, you still have to go over every line of your work to correct it, but the total time for transcribing interviews can be reduced by half or more. (More about VR software in appendix E) (**Further Reading:** transcription).

Recording Is Not a Substitute for Taking Notes

Finally, never substitute recording for note taking. Take notes during the interview *about* the interview. Did the informant seem nervous or evasive? Were there a lot of interruptions? What were the physical surroundings like? How much probing did you have to do? Take notes on the contents of the interview, even though you get every word on the machine.

A few informants will let you use a recorder but will balk at your taking notes. Don't assume, however, that informants will be offended if you take notes. Ask them. Most of the time, all you do by avoiding note taking is lose a lot of data. Informants are under no illusions about what you're doing. You're interviewing them. You might as well take notes and get people used to it, if you can.

FOCUS GROUPS

Focus groups are recruited to discuss a particular topic—anything from people's feelings about brands of beer to their experience in toilet training their children. The method derives from work by Paul Lazarsfeld and Robert Merton in 1941 at Columbia University's Office of Radio Research. A group of people listened to a recorded radio program that was supposed to raise public morale prior to America's entry into World War II. The listeners were told to push a red button whenever they heard something that made them react negatively and to push a green button when they heard something that made them react positively. The reactions were recorded automatically by a primitive polygraph-like apparatus. When the program was over, an interviewer talked to the group of listeners to find out why they had felt positively or negatively about each message they'd reacted to (Merton 1987) (box 8.4).

BOX 8.4

NOT ALL GROUP INTERVIEWS ARE *FOCUS GROUP* INTERVIEWS

Sometimes, you just find yourself in an interview situation with a lot of people. You're interviewing someone and other people just come up and insert themselves into the conversation. This happens spontaneously all the time in long-term fieldwork in small communities, where people all know one another. Rachel Baker (1996a, 1996b), for example, studied homeless boys in Kathmandu. When she interviewed boys in temples or junkyards, others might come by and be welcomed into the conversation-interview situation. If you insist on privacy, you might find yourself with no interview at all. Better to take advantage of the situation and just let the information flow. Just be sure to take notes on who's there, who's dominant, who's just listening, and so on, in any group interview.

The commercial potential of Lazarsfeld and Merton's pioneering work was immediately clear. The method of real-time recording of people's reactions, combined with focused interviewing of a group, is today a mainstay in advertising research and product design. MCI, the now defunct long-distance phone company, used focus groups to develop their advertising when they were just starting out. They found that customers didn't blame AT&T for the high cost of their long-distance phone bills; they blamed themselves for talking too long on long-distance calls. MCI came out with the advertising slogan: "You're not talking too much, just spending too much." The rest, as they say, is history (Krueger 1994:33).

Why Are Focus Groups So Popular?

The focus group method was a commercial success from the 1950s on, but it lay dormant in academic circles for more than 20 years. This is probably because the method is virtually devoid of statistics. Beginning in the late 1970s, however, interest among social researchers of all kinds boomed as researchers came to understand the benefits of combining qualitative and quantitative methods.

Focus groups do not replace surveys, but rather complement them. You can convene a focus group to discuss questions for a survey. Do the questions seem arrogant to respondents? Appropriate? Naive? A focus group can discuss the wording of a particular question or offer advice on how the whole questionnaire comes off to respondents. And you can convene a focus group to help interpret the results of a survey. But focus groups are not just adjuncts to surveys. They are widely used to find out *why* people feel as they do about something or the steps people go through in making decisions.

Two Cases of Focus Groups

Knodel et al. (1984), for example, used focus groups to study the fertility transition in Thailand. They held separate group sessions for married men under 35 and married women under 30 who wanted three or fewer children. They also held separate sessions for men and women over 50 who had at least five children. This gave them four separate groups. In all cases, the participants had no more than an elementary school education.

Knodel et al. repeated this four-group design in six parts of Thailand to cover the religious and ethnic diversity of the country. The focus of each group discussion was on the number of children people wanted and why.

Thailand was going through fertility transition in the 1980s, and the focus group study illuminated the reasons for the transition. "Time and again," these researchers report, "when participants were asked why the younger generation wants smaller families than the older generation had, they responded that nowadays everything is expensive" (Knodel et al. 1984:302).

People also said that all children, girls as well as boys, needed education to get the jobs that would pay for the more expensive, monetized lifestyle to which people were then becoming accustomed. It is, certainly, easier to pay for the education of fewer children. These consistent responses are what you'd expect in a society undergoing fertility transition.

Ruth Wilson et al. (1993) used focus groups in their study of acute respiratory illness (ARI) in Swaziland. They interviewed 33 individual mothers, 13 traditional healers, and 17 health care providers. They also ran 33 focus groups: 16 male groups and 17 female groups. The groups had from 4 to 15 participants, with an average of 7.

Each individual respondent and each group was presented with two hypothetical cases. Wilson et al. asked their respondents to diagnose each case and to suggest treatments. Here are the cases:

> Case 1. A mother has a 1-year-old baby girl with the following signs: coughing, fever, sore throat, running or blocked nose, and red or teary eyes. When you ask the mother, she tells you that the child can breast-feed well but is not actively playing.
>
> Case 2. A 10-month-old baby was brought to a health center with the following signs: rapid/difficult breathing, chest indrawing, fever for one day, sunken eyes, coughing for three days. The mother tells you that the child does not have diarrhea but has a poor appetite.

Many useful comparisons were possible with the data from this study. For example, mothers attributed the illness in Case 2 mostly to the weather, heredity, or the child's home environment. The male focus groups diagnosed the child in Case 2 as having asthma, fever, indigestion, malnutrition, or worms.

Wilson et al. (1993) acknowledge that a large number of individual interviews make it easier to estimate the degree of error in a set of interviews. However, they conclude that the focus groups provided valid data on the terminology and practices related to ARI in Swaziland. Wilson and her coworkers did, after all, have 240 respondents in their focus groups; they had data from in-depth interviews of all categories of persons involved in treating children's ARI; and they had plenty of participant observation in Swaziland to back them up.

Note some very important things about these studies. First, neither of them was based on a *single focus group* but on *a series of focus groups*. Second, in both studies, the groups were homogeneous with respect to certain independent variables—gender, number of children desired or produced, ethnicity—just as we saw with respect to experimental and sampling design. Finally, in the study by Knodel et al., the 24 groups were chosen to represent a subgroup in a factorial design—again just as we saw with experiments in chapter 4 and with sampling design in chapter 5. In other words, these focus group studies were designed to provide not only in-depth data about the reasons behind people's behavior, but data that could be systematically compared across groups.

Are Focus Groups Valid?

Ward et al. (1991) compared focus group and survey data from three studies of voluntary sterilization (tubal ligation or vasectomy) in Guatemala, Honduras, and Zaire. Ward et al. report that, "Overall, for 28% of the variables the results were similar" in the focus group and survey data. "For 42% the results were similar but focus groups provided additional detail; for 17% the results were similar, but the survey provided more detail. And in only 12% of the variables were the results dissimilar" (p. 273).

In the Guatemala study, 97% of the women surveyed reported no regrets with their decision to have a tubal ligation. The "vast majority" of women in the focus groups also reported no regrets. This was counted as a "similar result." Ten percent of the women surveyed reported having had a tubal ligation for health reasons. In the focus groups, too, just a few women reported health factors in their decision to have the operation, but they provided more detail and context, citing such things as complications from previous pregnancies.

This is an example of where the focus group and survey provide similar results, but where the focus group offers more detail. Data from the focus groups and the survey confirm that women heard about the operation from similar sources, but the survey shows that 40% of the women heard about it from a sterilized woman, 26% heard about it from a health professional, and so on. Here, the survey provides more detail, though both methods produce similar conclusions.

Gillespie (1992) compared the results of four focus groups with data from a survey of the same population and found that results were dissimilar in 18% of the variables. The bottom line: Focus groups—like participant observation, in-depth interviews, and other systematic qualitative methods—should be used for the collection of data about content and process and not for estimating population parameters of personal attributes. The belief that a woman has or does not have a right to an abortion is a personal attribute, like gender, age, annual income, or religion. If you want to estimate the proportion of

people in a population who believe that a woman has a right to an abortion, then focus groups are not the method of choice.

A proportion is a number, and if you want a good number—a valid one, a useful one—you need a method that produces exactly that. A survey, based on a representative sample, is the method of choice here. But if you want information about content—about *why* people think a woman should or should not have the right to an abortion—then that's just the sort of thing a focus group can illuminate.

RUNNING A FOCUS GROUP

The group moderator gets people talking about whatever issue is under discussion. Leading a focus group requires the combined skills of an ethnographer, a survey researcher, and a therapist. You have to watch out for people who want to show off and close them down without coming on too strongly. You have to watch out for shy people and draw them out, without being intimidating. Tips on how to do all this, and a lot more, are in *The Focus Group Kit*, a series of six how-to books (D. L. Morgan and Krueger 1998). Don't even think about getting into focus group management without going through this kit (box 8.5).

BOX 8.5

COMPOSITION OF A FOCUS GROUP

Focus groups typically have 6–12 members, plus a moderator. Seven or 8 people is a popular size. If a group is too small, it can be dominated by 1 or 2 loudmouths; if it gets beyond 10 or 12, it gets tough to manage. Smaller groups are better when you're trying to get really in-depth discussions going about sensitive issues (D. L. Morgan 1997). Of course, this assumes that the group is run by a skilled moderator who knows how to get people to open up and how keep them opened up.

The participants in a focus group should be more or less homogeneous and, in general, should not know one another. Richard Krueger, a very experienced focus group moderator, says that "familiarity tends to inhibit disclosure" (1994:18). It's easy to open up more when you get into a discussion with people whom you are unlikely ever to see again (sort of like what happens on long air flights). Obviously, what "homogeneous" means depends on what you're trying to learn. If you want to know why a smaller percentage of middle-class African American women over 40 get mammograms than do their white counterparts, then you need a group of middle-class African American women who are over 40.

In a focus group about sensitive issues like abortion or drug use, the leader works at getting the group to gel and getting members to feel that they are part of an understanding cohort of people. If the group is run by an accomplished leader, one or more members will eventually feel comfortable about divulging sensitive information about themselves. Once the ice is broken, others will feel less threatened and will join in. Moderators should not be known to the members of a focus group, and focus group members should not be employees of a moderator. Hierarchy is not conducive to openness.

In running a focus group, remember that people will disclose more in groups that are supportive and nonjudgmental. Tell people that there are no right or wrong answers to the questions you will ask and emphasize that you've invited people who are similar in their backgrounds and social characteristics. This, too, helps people open up (Krueger 1994:113).

Above all, don't lead too much and don't put words in people's mouths. In studying nutritional habits, don't ask a focus group why they eat or don't eat certain foods; do ask them to talk about what kinds of foods they like and dislike and why. In studying risky sexual behavior, don't ask, "Do you use condoms whenever you visit a prostitute?"; do ask people to talk about their experience with prostitutes and exactly what kind of sexual practices they prefer. Your job is to keep the discussion on the topic. Eventually, people will hit on the nutritional habits or the sexual acts that interest you, and you can pick up the thread from there.

Analyzing Data from Focus Groups

You can analyze focus group data with the same techniques you would use on any corpus of text: field notes, life histories, open-ended interviews, and so on. As with all large chunks of text, you have two choices for very different kinds of analysis. You can do formal content analysis, or you can do qualitative analysis. See chapters 18 and 19 for more about this.

As with in-depth interviews, it's best to record (or videotape) focus groups. This is a bit tricky, though, because any audio of a focus group, whether digital or tape, is hard to understand and transcribe if two or more people talk at once. A good moderator keeps people talking one at a time. Don't hide the recorder or the microphones. Someone is sure to ask if they're being recorded, and when you tell them, "Yes"—which you must do—they're sure to wonder why they had to ask.

If you are just trying to confirm some ideas or to get a general notion of how people feel about a topic, you can simply take notes from the audio and work with your notes. Most focus groups, however, are transcribed. The real power of focus groups is that they produce ethnographically rich data. Only transcription captures a significant part of that richness. But be prepared to work with a lot of information. Any single hour-and-a-half focus group can easily produce 50 pages or more of text.

Many focus groups have two staff members: a moderator and a person who does nothing but jot down the name of each person who speaks and the first few words they say. This makes it easier for a transcriber to identify the voices. If you can't afford this, or if you feel that people would be uncomfortable with someone taking down their names, you can call on people by name, or mention their name when you respond to them. Things can get rolling in a focus group (that's what you want), and you'll have a tough time transcribing the audio if you don't know who's talking (**Further Reading:** focus groups).

RESPONSE EFFECTS

Response effects are measurable differences in the responses of people being interviewed that are predictable from characteristics of the interviewers and those being interviewed—like whether the sex or race or age of interviewer and of the respondent are the same or different—and dozens of other things (box 8.6).

As early as 1929, Stuart Rice showed that the political orientation of interviewers can effect what they report people told them. Rice was doing a study of derelicts in flop houses

BOX 8.6

THE EXPECTANCY EFFECT

In 1966, Robert Rosenthal and Lenore Jacobson (1968) conducted an experiment. At the beginning of the school year, they told some teachers at a school that the children they were about to get had tested out as "spurters." That is, according to tests, they said, those particular children were expected to make significant gains in their academic scores during the coming year. Sure enough, those children did improve dramatically—which was really interesting, because Rosenthal and Jacobson had matched the "spurter" children and teachers at random.

This experiment showed the power of the **expectancy effect**, or "the tendency for experimenters to obtain results they expect, not simply because they have correctly anticipated nature's response but rather because they have helped to shape that response through their expectations" (Rosenthal and Rubin 1978:377).

Strictly speaking, the expectancy effect is not a response effect at all. But for fieldworkers, it is an important effect to keep in mind. If you are studying a small community, or a neighborhood in a city, or a hospital or clinic for a year or more, interacting daily with a few key informants, your own behavior can affect theirs in subtle (and not so subtle) ways, and vice versa. Don't be surprised if you find your own behavior changing over time in relation to key informants.

and he noticed that the men contacted by one interviewer consistently said that their down-and-out status was the result of alcohol; the men contacted by the other interviewer blamed social and economic conditions and lack of jobs. It turned out that the first interviewer was a prohibitionist and the second was a socialist (cited in Cannell and Kahn 1968:549). Katz (1942) found that middle-class interviewers got more politically conservative answers in general from lower-class respondents than did lower-class interviewers, and Robinson and Rhode (1946) found that interviewers who looked non-Jewish and had non-Jewish-sounding names were almost *four times more likely* to get anti-Semitic answers to questions about Jews than were interviewers who were Jewish looking and who had Jewish-sounding names.

Since these pioneering efforts, hundreds of studies have been conducted on the impact of things like race, sex, age, and accent of both the interviewer and the informant; features of the environment where the interview takes place (like whether the interview is done in private or in the presence of a third party); the nature of the task that people are asked to perform (like whether the respondent is asked to write out an answer, in text, or to just circle a number on a form); the mode of the interview (like comparing face-to-face, telephone, and Internet interviews about the same topic).

Sex-of-interviewer effects have been the focus of many studies. Hyman and Cobb (1975), for example, found that female interviewers who took their cars in for repairs themselves (as opposed to having their husbands do it) were more likely to have female respondents who reported getting their own cars repaired. Zehner (1970) found that when women in the United States were asked by women interviewers about premarital

sex, they were more inhibited than if they were asked by men. Male respondents' answers were not affected by the gender of the interviewer. McCombie and Anarfi (2002) found the same sex-of-interviewer effect 30 years later in Ghana: Young men (15–18 year olds) were equally likely to tell male or female interviewers that they had had sex, but young women were more likely to divulge this to male interviewers than to female interviewers. In the Tamang Family Research Project in Nepal, William Axinn (1991) found that women were simply better than men as interviewers: The female interviewers had significantly fewer "don't know" responses than did the male interviewers. Axinn supposes this might be because the survey dealt with marital and fertility histories. In a multi-year study in Kenya of women's networks and their AIDS-related behavior, Alex Weinreb (2006) found that the most reliable data were collected by female-*insider* interviewers—that is, women from the local area who were trained to be interviewers for the project—compared to the *stranger*-interviewers who were brought in from the outside (box 8.7).

The Deference Effect

When people tell you what they think you want to know, so as not to offend you, that's called the **deference effect**. Aunger (1992, 2004) may have experienced this in Zaire (see box 8.7). In fact, it happens all the time, and researchers have long been aware of the problem. In 1958, Lenski and Leggett embedded two contradictory questions in a face-to-face interview, half an hour apart. Respondents were asked whether they agreed or disagreed with the following two statements: (1) It's hardly fair to bring children into the world, the way things look for the future; (2) Children born today have a wonderful future to look forward to. Just 5% of Whites agreed with *both* statements compared to 20% of African Americans. Lenski and Leggett concluded that this was the deference effect in action: Blacks were four times more likely than Whites to agree to anything, even contradictory statements, because the interviewers were almost all white and of higher perceived status than the respondents (Lenski and Leggett 1960).

In the National Black Election Study, 872 African Americans were polled before and after the 1984 presidential election. Since interviewers were assigned randomly to respondents, some people were interviewed by a white person before the election and an African American after the election. And vice versa: Some people were interviewed by an African American before the election and a white person on the second wave. When African American interviewers in the preelection polls were replaced by white interviewers in the postelection surveys, African Americans were more likely to say that Blacks don't have the power to change things, that Blacks can't make a difference in local or national elections, that Blacks cannot form their own political party, and that Whites are not responsible for keeping Blacks down—very powerful evidence of a race-of-interviewer effect (D. W. Davis 1997) (box 8.8).

Reese et al. (1986:563) tested the deference effect in a telephone survey of Anglo and Mexican American respondents. When asked specifically about their cultural preference, 58% of Hispanic respondents said they preferred Mexican American culture over other cultures, irrespective of whether the interviewer was Anglo or Hispanic. Just 9% of Anglo respondents said they preferred Mexican American culture when asked by Anglo interviewers, but 23% said they preferred Mexican American culture when asked by Hispanic interviewers.

Questions about gender and gender roles produce deference effects, too. When you ask people in the United States how most couples actually divide child care, men are more likely than women to say that men and women share this responsibility—if the interviewer is a man (Kane and McCaulay 1993:11). Do women have too much influence, just

BOX 8.7

THE INSIDER-INTERVIEW EFFECT IN THE ITURI FOREST

Robert Aunger (1992, 2004:145–62) studied three groups of people in the Ituri forest of Zaire. The Lese and Budu are horticultural, and the Efe are foragers. Aunger wanted to know if they shared the same food avoidances. He and three assistants, two Lese men and one Budu man, interviewed a total of 65 people. Each of the respondents was interviewed twice and was asked the same 140 questions about a list of foods.

Aunger identified two types of errors in his data: forgetting and mistakes. If informants said in the first interview that they did not avoid a particular food but said in the second interview that they did avoid the food, Aunger counted the error as forgetfulness. If informants reported in interview two a different type of avoidance for a food than they'd reported in interview one, then Aunger counted this as a mistake.

Even with some missing data, Aunger had over 8,000 pairs of responses in his data (65 *pairs* of interviews, each with up to 140 responses), so he was able to look for the causes of discrepancies between interview one and interview two. About 67% of the forgetfulness errors and about 79% of the mistake errors were correlated with characteristics of informants (gender, ethnic group, age, and so on). However, about a quarter of the variability in what informants answered to the same question at two different times was due to characteristics of the interviewers (ethnic group, gender, native language, etc.), and about 12% of variability in forgetting was explained by interviewer experience. As the interviewers interviewed more and more informants, the informants were less likely to report "no avoidance" on interview one and some avoidance on interview two for a specific food. In other words, interviewers got better and better with practice at drawing out informants on their food avoidances.

Of the four interviewers, though, the two Lese and the Budu got much better, while the anthropologist made very little progress. Was this because of Aunger's interviewing style, or because informants generally told the anthropologist different things than they told local interviewers, or because there is something special about informants in the Ituri forest? We'll know when we add variables to Aunger's study and repeat it in many cultures, including our own (**Further Reading:** response effects).

the right amount of influence, or too little influence in today's society? When asked *this* question by a male interviewer, men are more likely to say that women have *too much* influence; when asked the same question by a female interviewer, men are more likely to say that women have *too little* influence. And similarly for women: When asked by a female interviewer, women are more likely to say that men have *too much* influence than when asked by a male interviewer (Kane and Macaulay 1993:14–15).

Lueptow et al. (1990) found that women gave more liberal responses to female interviewers than to male interviewers on questions about gender roles. Men's attitudes about gender roles were, for the most part, unaffected by the gender of the interviewer—except

BOX 8.8

BARACK OBAMA AND THE BRADLEY EFFECT

In 1982, Tom Bradley, the mayor of Los Angeles, ran against George Deukmejian for the office of governor of California. Bradley was ahead in the polls for the governorship of California right up to election day—and lost. Some voters had told pollsters that they were for Bradley, who is black, and then voted for Deukejian, who is white. The so-called Bradley effect was at work in 1989, when Douglas Wilder, an African American, ran against Marshall Coleman, who is white, for the governorship of Virginia. Preelection polls showed that Wilder was far ahead, but in the end, he won by only a slim margin. White voters were more likely to claim Wilder as their choice if the interviewer was African American than if the interviewer was white (Finkel et al. 1991). Barack Obama is widely credited with ending the Bradley Effect, but he lost the 2008 New Hampshire primary to Hillary Clinton by three points after being ahead in the polls by eight points—right up to election day (Kohut 2008).

■

that highly educated men gave the *most* liberal responses about gender roles to female interviewers. "It appears," said Lueptow et al., "that educated respondents of both sexes are shifting their answers toward the socially desirable positions they think are held by female interviewers" (p. 38). Attitudes about gender roles sure are adaptable. That was in 1990. In 2008, about 26% of the American public was "angry or upset" at the prospect of a woman president, even though, at the time almost 90% of Americans told pollsters that they would vote for a qualified woman foir president (Streb et al. 2008:77).

Questions that aren't race related, by the way, are not affected much by the race or the ethnicity of either the interviewer or the respondent. Still, whenever you have multiple interviewers, keep track of the race, ethnicity, and gender of the interviewer and test for response effects. Identifying sources of bias is better than not identifying them, even if you can't eliminate them (**Further Reading:** the deference effect).

The Social Desirability Effect

When people tell you what they think will make them look good, especially according to prevailing standards of behavior and thought, that's the social desirability effect. Hadaway et al. (1998) went to a large Protestant church and found 115 people in attendance at the Sunday school. On Monday morning, when Hadaway et al. polled the whole church membership, 181 people claimed to have been in Sunday school the previous day. Headcount experiments like this one typically produce estimates of church attendance that are 55%–59% of what people report (T. W. Smith 1998).

The social desirability effect is influenced by the way you ask the question. Major surveys, like the Gallup Poll, ask something like: "How often do you attend religious services?" Then they give the people choices like "once a week, once a month, seldom, never." Presser and Stinson (1998) asked people on Monday to list everything they had done did from "midnight Saturday to midnight last night." When they asked the question this way, 29% of respondents said that they had gone to church. Asking "How often do you go to church?" produced estimates of 37%–45%. This is a 28%–50% *difference* in reported behavior and is statistically very significant (**Further Reading:** social desirability effect).

The Third-Party-Present Effect

We sort of take it for granted that interviews are private conversations, conducted one on one, but in fact, many face-to-face interviews have at least one third party in the room, often the spouse or partner of the person being interviewed. Does this affect how people respond to questions? As with other response effects, the answer is yes, sometimes. Zipp and Toth (2002), for example, analyzed data from a household survey in Britain and found that when the spouses are interviewed together, they are much more likely to agree about many things—like who does what around the house—than when they are interviewed separately. Apparently, people listen to each other's answers and modify their own answers accordingly, which puts on a nice, unified face about their relationship.

As you'd expect, there is a social desirablity effect when a third party is present. Casterline and Chidambaram (1984) examined data from 24 developing countries in the World Fertility Study and found that women in those countries are less likely to admit using contraception when a third party is present at the interview. Anthropologists face this situation a lot: trying to get people to talk about sensitive topics and assuring them of privacy, but unable to find the privacy for an interview.

On the other hand, Aquilino (1993) found that when their spouse is in the room, people report more marital conflict than when they are interviewed alone. They are also more likely to report that they and their spouse lived together before marriage if their spouse is in the room. Perhaps, as Mitchell (1965) suggested 45 years ago, people own up more to sensitive things like this when they know it will be obvious to their spouse that they are lying. Seems like a good thing to test (**Further Reading:** third-party-present effect).

Threatening Questions

In general, if you are asking someone a nonthreatening question, like whether they have a library card, then response effects are minimal. But if you ask people about their alcohol consumption, or whether they ever shoplifted when they were children, or whether they have family members who have had mental illness, or how many sexual partners they've had, then response effects are really important. One key finding on this problem is, intuitively, that disclosure of information about socially undesirable behavior increases with the perception people have of their anonymity (Tourangeau and Yan 2007). So, people open up more on questionnaires about illegal or embarrassing behavior that are self-administered than in surveys conducted face-to-face, and still more when they think that a survey is truly anonymous. (See the section on the randomized response technique and the section on computerized interviews in chapter 9).

Asking about other people increases reports about socially undesirable behavior. Katz and Naré (2002) asked 1,973 single Muslim women between the ages of 15 and 24 in Dakar, Senegal, if they had ever been pregnant. Three percent of the women said they had. But 25% of the same women said that at least one of their *three closest friends* had been pregnant—more than eight times what they reported about themselves. (See Sudman et al. [1977:147–51] on the three-closest-friends technique.)

Asking the interviewers on a project to record their interviews produces a higher response rate, particularly to sensitive questions about things like sexual behavior. Apparently, when interviewers know that their work can be scrutinized (from the recordings), they probe more and get informants to open up more (Billiet and Loosveldt 1988).

And if you give people choices that include a big number of any behavior, you'll probably get reports of more of that behavior. Tourangeau and Smith (1996) asked men and women: "During the last 12 months, that is, since August/September 1993, how many

men [women], if any, have you had intercourse with?" Some people were asked simply to tell the interviewer a number. Others were asked to choose one of the following: 0, 1, 2, 3, 4, 5 or more. And still others were asked to choose one of the following: 1–4, 5–9, 10–49, 50–99, 100 or more. People reported more sex partners when given high-end choices than when given low-end choices or an open-ended question (Tourangeau and Smith 1996:292). In Germany, people reported watching more television when they were given choices that included a big number of hours (Schwarz et al. 1985).

You might be surprised, though, at what counts as a threatening question. R. A. Peterson (1984) asked 1,324 people one of the following questions: (1) How old are you? (2) What is your age? (3) In what year were you born? or (4) Are you 18–24 years of age, 25–34, 35–49, 50–64, 65 or older? Then Peterson got the true ages for all the respondents from reliable records. There was no significant difference in the accuracy of the answers obtained with the four questions, but almost 10% of respondents refused to answer question 1, while only 1% refused to answer question 4, and this difference *is* significant (**Further Reading:** asking threatening questions).

ACCURACY

Even when people tell you what they think is the absolute truth, there is still the question of whether the information they give you is accurate.

A lot of research—ethnographic and survey research alike—is about mapping opinions and attitudes. When people tell you that they *approve of* how the chief is handling negotiations for their village's resettlement, or when they tell you that they *prefer* a particular brand of beer to some other brand, they're talking about internal states. You pretty much have to take their word for such things.

But when we ask people to tell us about their actual behavior (How many times did you take your baby to the clinic last month? How many times last year did you visit your mother's village?), or about their environmental circumstances (How many hectares of land do you have in maize? How many meters is it from your house to the well?), we can't just assume informant accuracy (box 8.9).

We see reports of behavior in our local newspapers all the time: College students today are binge drinking more than they did 5 years ago. Americans are going to church less often than they did a decade ago.

In back of *findings* like these are *questions* like these:

Circle one answer:
How many times last month did you consume five or more beers or other alcoholic drinks in a single day?

Never
Once
Twice
Three times
More than three times

How often do you go to church?

Never
Occasionally—once a month or less
About once a week
More than once a week

BOX 8.9

DIETARY RECALL

Studies of diet and human nutrition mostly rely on informants to recall what they've eaten over the past 24 hours or what they usually eat for various meals. They often produce dreadfully inaccurate results.

C. J. Smith et al. (1996) compared the responses of 575 Pima and Papago Indians (in Arizona) to a 24-hour recall instrument about food intake with responses to a very detailed survey called the Quantitative Food Frequency questionnaire. In the QFF, interviewers probe for a list of regularly consumed foods in a community. Smith et al. also assessed the energy expenditure of 21 people in the research group using the doubly labeled water technique. The DLW technique involves giving people special water to drink—water with isotopes that can be tracked in blood and urine samples—and then testing, over time, their actual intake of nutrients.

The correlation, across the 21 participants, between the energy intake measured by the DLW technique and the energy intake estimated by the informants' responses to the QFF, was 0.48. This correlation is statistically significant, but it means that just 23% (0.48^2) of the variation in actual energy intake across the 21 people was accounted for by their responses to a very detailed interview about their food consumption. And the correlation of actual energy intake with estimates from the 24-hour recall data was much worse.

R. K. Johnson et al. (1996) also found no useful relation between individual 24-hour recall measurements of energy intake among children in Vermont and measurements of those same children by the DLW technique. But, in the all-is-not-lost department, Johnson et al. found that averaging the data for energy intake across *three* 24-hour recalls in 14 days (on day 1, day 8, and day 14) produced results that were very similar to those produced by the DLW technique. So, people hover around giving accurate answers to a question about calorie intake and if you get at least three answers for three time windows and take the average, you may get a useful result (**Further Reading:** measuring food intake and physical activity).

La Pierre Discovers the Problem

We've known for a long time that we should be suspicious of this kind of data. From 1930 to 1932, Richard La Pierre, accompanied by a Chinese couple, crisscrossed the United States, twice, by car. The threesome covered about 10,000 miles, stopping at 184 restaurants and 66 hotels. And they kept records. There was a lot of prejudice against Chinese in those days, but they were not refused service in a single restaurant and just one hotel turned them away (La Pierre 1934).

Six months after the experiment ended, La Pierre sent a questionnaire to each of the 250 establishments where the group had stopped. One of the things he asked was: "Will you accept members of the Chinese race as guests?" Ninety-two percent—230 out of 250—replied "No."

By today's standards, La Pierre's experiment was crude. He could have surveyed a control group—a second set of 250 establishments that they hadn't patronized but that

were in the same towns where they'd stopped. With self-administered questionnaires, he couldn't be sure that the people who answered the survey (and who claimed that they wouldn't serve Chinese) were the same ones who had actually served the threesome. And La Pierre didn't mention in his survey that the Chinese couple would be accompanied by a white man.

Still, La Pierre's experiment was terrific for its time. It made clear that what people say they do (or would do) is not a proxy for what they actually do or will do (see Deutscher 1973). This basic finding shows up in what you might think were the most unlikely places: In the 1961 census of Addis Ababa, Ethiopia, 23% of the women underreported the *number of their children*. Apparently, people there didn't count babies who die before reaching the age of 2 (Pausewang 1973:65). People in the United States often omit newborns when they fill out the Decennial Census form (Dillman et al. 2009b:225), and in China today, if a child dies soon after birth, couples may decide to report neither the birth nor the death and instead try to conceive again as quickly as possible. And, under the one-child policy, the births of female babies may not be reported at all because of the desire by couples to have a son (Merli and Rafferty 2000:110).

Why People Are Inaccurate Reporters of Their Own Behavior

People are inaccurate reporters of their own behavior for many reasons. Here are four:

1. Once people agree to be interviewed, they have a personal stake in the process and usually try to answer all your questions—whether they understand what you're after or not.
2. Human memory is fragile, although it's clearly easier to remember some things than others.

Cannell et al. (1961) found that the ability to remember a stay in the hospital is related to the length of the stay, the severity of the illness that lands you there, and whether or not surgery is involved. It's also strongly related to the length of time since discharge. Cannell and Fowler (1965) found that people report accurately 90% of all overnight hospital stays that happened 6 months or less before being interviewed.

It's easy for people to remember a rare event, like surgery, that occurred recently. But, as Sudman and Schwarz (1989) point out, if you ask people to think about some common behavior going back months at a time, they probably use estimation rules. When Sudman and Schwartz asked people "How many [sticks] [cans] of deodorant did you buy in the last 6 months?" they started thinking: "Well, I usually buy deodorant about twice a month in the summer, and about once a month the rest of the year. It's now October, so I suppose I must have bought 10 deodorants over the last 6 months." And then they say, "10," and that's what you write down.

3. Interviews are social encounters. People manipulate those encounters to whatever they think is their advantage.

Adolescent boys tend to exaggerate, and adolescent girls tend to minimize, reports of their own sexual experience (see Catania et al. 1996).

4. People can't count a lot of behaviors, so they use rules of inference.

In some situations, they invoke D'Andrade's "what goes with what" rule (1974) and report what they *suppose* must have happened, rather than what they actually saw. Free-

man et al. (1987) asked people in their department to report on who attended a particular colloquium. People who were *usually* at the department colloquium were mentioned as having attended the particular colloquium—even by those who hadn't attended (and see Shweder and D'Andrade 1980).

Reducing Errors: Jogging Informants' Memories

Loftus and Marburger (1983) found that landmarks help reduce forward telescoping—where people report that something happened 1 month ago when it really happened 2 months ago (backward telescoping is rare). The title of their article says it all: "Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events." Means et al. (1989) asked people to recall landmark events in their lives going back 18 months from the time of the interview.

Once the list of personal landmark events was established, people were better able to recall hospitalizations and other health-related events. In the field, as you do life history interviews, try to establish personal milestones for each informant—like their first hunting kill or their clitoridectomy or, for older informants, burying their parents or becoming grandparents—and ask them to report on what has happened since each landmark.

Aided recall increases the number of events recalled, but also appears to increase the telescoping effect (Bradburn 1983:309). In studies where you interview people more than once, you can correct for telescoping by reminding them what they said last time in answer to a question and then asking them about their behavior since their last report.

Event history and **life history calendars** are effective aids to recall and are particularly useful in societies where there are no written records. Leslie et al. (1999:375–78), for example, developed an event calendar for the Ngisonyoka section of the South Turkana pastoralists in northwestern Kenya. The Turkana name their seasons rather than their years. Based on many interviews between 1983 and 1984, Leslie et al. were able to build up a list of 143 major events associated with seasons between 1905 and 1992. Events include things like "no hump" in 1961 (it was so dry that the camels' humps shrank), "bulls" in 1942 (when their bulls were taken to pay a poll tax), and "rescue" in 1978 (when rains came). This painstaking work has made it possible for many researchers to gather demographic and other life history data from the Ngisonyoka Turkana. William Axinn and colleagues (1999:252) also used multiple event cues in Nepal and, like Leslie et al., report that this was particularly helpful for older informants (**Further Reading:** event- and life-history calendars).

If you are working in an industrialized environment with literate informant, you can ask people to review their credit card statements and long-distance phone bills and to remember events, places, and people associated with each credit or phone event. College transcripts help people think about what they were doing and the people they met along the way. Still . . . Horn (1960) asked people to report their bank balance. Of those who did not consult their bankbooks, just 31% reported correctly. Those who consulted their records didn't do that much better. Only 47% reported correctly (reported in Bradburn 1983:309).

It's different, of course, in nonindustrialized societies. When Elliot Fratkin (2004:19) asked Ariaal warriors in northern Kenya how many cattle they owned, the answer was always "many." Veterinary studies showed that the herds of the Samburu (a group closely related to the Ariaal) were two-thirds female and that 50% of those cattle were lactating at any time. A household with 6 nursing calves, Fratkin calculated, would have, on average, 12 cows and 4 female calves, plus 8 male cattle.

Informant accuracy remains a major problem. Gary Wells and colleagues (2003)

showed a video of a staged crime to 253 students. Then they showed the students a photo lineup of six people and asked the students to pick out the culprit. Every single student picked one of the six photos, but there was a small problem: The culprit wasn't in the six photos. We need a lot more research about the rules of inference that people use when they respond to questions about where they've been, who they were with, and what they were doing (**Further Reading:** informant accuracy).

FURTHER READING

Transcription: Bailey (2008); Bucholtz (2000); Du Bois (1991); Duranti (2006); Maloney and Paolisso (2001); Matheson (2007); McLellan et al. (2003); Ochs (1979).

Interviewing: Graham (2000); Gubrium and Holstein (2002); Kvale and Brinkman (2009); McCracken (1988); Mishler (1986); Rubin (2005); Seidman (2006); Wengraf (2001).

Focus groups: Krueger (1994); Morgan (1997); Morgan and Krueger (1998); Sayles et al. (2007); Stewart and Shamdasani (1990); Vaughn et al. (1996).

Response effects: Borgers et al. (2004); Bradburn (1983); Cannell et al. (1979); Christian et al. (2009); Javeline (1999); Johnson and Delameter (1976); Schaeffer and Presser (2003); Schuman and Presser (1979, 1981); Schwarz (1999); Schwarz et al. (1991); Singer et al. (1983); Sudman and Bradburn (1974); Tourangeau et al. (2000).

Deference effect: Dotinga et al. (2005); Krysan and Couper (2003).

Social desirability effect: DeMaio (1984); Kreuter et al. (2008); Phillips and Clancy (1972); Press and Townsley (1998); W. W. Smith (2006); van de Mortel (2008).

Third-party present effect: Aquilino (1997); Aquilino et al. (2000); Blair (1979); Boeije (2004); Bradburn (1979); Hartmann (1994); Pollner and Adams (1997); T. W. Smith (1997).

Asking threatening questions: Bradburn et al. (1978); Bradburn, Sudman et al. (1979); Catania et al. (1996); Gribble et al. (1999); Hewitt (2002); Johnston and Walton (1995); Makkai and McAllister (1992); Mooney and Gramling (1991); Wiederman et al. (1994).

Event- and life-history calendars: Belli (1998); Caspi et al. (1996); Freedman et al. (1988); Glassner and van der Vaart (2009); Kessler and Wethington (1991); Martyn and Belli (2002); Yoshihama et al. (2005).

Measuring food intake and physical activity: Graham (2003); Hebert et al. (2002); Hill and Davies (2001); Leenders et al. (2000); Pelto et al. (1989); Prince et al. (2008); Quandt and Rittenbaugh (1986); Schoenberg (1997, 2000); Subar et al. (2003).

Informant accuracy: see **Further Reading**, chapter 2.

10

Interviewing III: Cultural Domains

Cultural domain analysis is the late-model version of **ethnoscience**, a movement in anthropology of the 1950s and 1960s (Sturtevant 1964). The goal of ethnoscience was to understand cultural systems of classification—that is, how people in a group think about lists of things that somehow go together. These can be lists of physical, observable things—plants, colors, animals, symptoms of illness—or conceptual things—occupations, roles, emotions. (For seminal work on modern cultural domain analysis, see Borgatti 1993/1994, 1999 and Weller and Romney 1988).

The spectrum of colors, for example, has a single physical reality that you can see on a machine. Some peoples across the world, however—Xhosa, Navajo, Ñähñu—identify the colors across the physical spectrum of green and blue with a single gloss. In Ñähñu, for example, the word is *nk'ami* and in Navajo it's *dootl'izh*. Linguists and cognitive scientists who study this phenomenon call this color "grue" (see, e.g., Davies et al. 1994; Gammack and Denby 2006; and Kim 1985).

This does *not* mean that people who have a word for grue fail to *see* the difference between things that are the color of grass and things that are the color of a clear sky. They just *label* chunks of the physical spectrum of colors differently than we do and use adjectival modifiers of grue to express color differences within the blue-green spectrum. In Navajo, turquoise is *yáago dootl'izh*, or "sky grue," and green is *tádlidgo dootl'izh*, or "water skum grue" (Oswald Werner, personal communication). If this seems exotic to you, get a chart of, say, 100 lipstick colors or house paint colors and ask people at your university to name the colors. On average, women will probably recognize (and name) more colors than men will; and art majors of both sexes will name more colors than, say, engineering majors will.

KINSHIP AND OTHER DOMAINS

This concern for understanding cultural differences in how people cut the natural world goes a long way back in anthropology—all the way to the early interest in kinship. Lewis Henry Morgan (1997 [1870]) studied systems of kinship nomenclature. His work made clear that if someone says, "This is my sister," you can't assume that they have the same mother and father. Lots of different people can be called "sister," depending on the kinship system. And in his work with the Murray Islanders (in the Torres Straits between Australia and Papua New Guinea) and then later with the Todas of southern India, W.H.R. Rivers developed the genealogical method—those ego-centered graphs for organizing kinship data that we take for granted today—as a way to elicit accurately and systematically the inventory of kin terms in a language (Rivers 1910, 1968 [1914]).

Anthropologists also noticed very early that, although kinship systems *could* be unique

to each culture—which would mean that each system required a separate set of rules—they simply weren't. Alfred Kroeber showed in 1909 that just eight features were needed to distinguish kinship terms in any system: (1) whether the speaker and the kin referred to were of the same or different generations; (2) the relative age people who are of the same generation—older or younger brother, for example; (3) whether the person referred to is a collateral or a lineal relative; (4) whether the person referred to is an affinal or consanguineal relative; (5) whether the relative is male or female; (6) whether the speaker is male or female; (7) whether the person who links the speaker and the relative is male or female; and (8) whether the person who links the speaker and the relative is alive or dead.

Now, if you first choose whether to use or not use any of those eight features and then choose among the two alternatives to each feature, you can concoct $3^8 = 6{,}561$ kinds of kinship systems. But, although there are some rare exceptions (the bilineal Yakö of Nigeria, the ambilineal Gilbert Islanders), most of the world's kinship systems are of one those familiar types you studied in Anthropology 101—the Hawaiian, Sudanese, Omaha, Eskimo, Crow, and Iroquois types. Early anthropologists found it pretty interesting that the world's real kinship systems comprised just a tiny set of the possibilities, and to this day, a small, hardy band of anthropologists continues to study the elements of these systems and how those elements are associated with particular political, economic, or environmental conditions (Kronenfeld 2009; White and Schweizer 1998) (**Further Reading:** kinship studies).

An interest in classifying kinship systems led to methods for discovering sets of terms in other domains, like kinds of foods, things to do on the weekend, kinds of crime, bad names for ethnic groups, dirty words, names for illnesses, etc. Note that none of these is about people's preferences. If we ask people which of two political candidates they favor in an election, we might also ask them about their income, their ethnicity, their age, and so on. Then we look for packages of variables about the people that predict their preference for a candidate. In cultural domain analysis, we're interested in the items that comprise the domain—the illnesses, the edible plants, the jobs that women and men do, etc.—and how those items are related to each other in people's minds (Borgatti 1999; Spradley 1979) (box 10.1). (More about building folk taxonomies in chapter 17.)

The methods for collecting data about the content and structure of cultural domains include *free lists*, *sentence frames*, *triad tests*, *pile sorts*, and *paired comparisons*. All of these methods produce a lot of data very quickly and some of them (particularly free lists and pile sorts) are even fun for people to do. And, with software, like ANTHROPAC (Borgatti 1992a) and UCINET (Borgatti et al. 2002), it's easy to analyze these data. We'll return to analyzing these kinds of data in chapter 16 (**Further Reading:** data collection for domain analysis).

FREE LISTING

Free listing is a deceptively simple, but powerful technique. Data from short, open-ended questions on surveys can be coded to produce lists, as can transcriptions of ethnographic interviews and focus groups. In free listing, however, we tell people: "List all the X you can think of," where X might be things they do on weekends, brands of cars, things people do when they get a cold, ways to avoid pregnancy, places in the community frequented by commercial sex workers, and so on.

The object is to get informants to list as many items as they can in a domain, so you need to probe and not just settle for whatever people say. Brewer et al. (2002:112) found that **semantic cueing** increased the recall of items in a free list by over 40%. Tell infor-

BOX 10.1

HOW SPRADLEY LEARNED TO NAVIGATE THE POLICE DEPARTMENT

Cultural domains are everywhere. Spradley (1979) reported that he once called the St. Paul, Minnesota, police department and said he needed to find the case number of a robbery that had been committed at his house. Two bicycles had been stolen from his garage in the middle of the night, while he was asleep. The police had investigated, but Spradley's insurance company needed the case number to process the claim. When Spradley told the police that he needed the case number for a "robbery," they quite naturally transferred his call to the robbery unit. But the people there couldn't help him because, according to their rules, robberies involve a face-to-face encounter between the criminal and the victim and the criminal uses a gun.

Spradley was transferred to burglary, but they couldn't help him either because, they said, theft of bicycles is handled by the juvenile division in St. Paul. Eventually, Spradley got his case number, but, he said, if he had understood the police culture, he "would have begun with a simple question: What part of the police department has records of bicycles stolen from a garage when no one is present?" (1979:142). In other words, if he'd known taxonomy for the cultural domain of crimes, he'd have asked the right question and gotten taken care of right away.

mants to: "Think of all the kinds of X [the domain] that are like Y," where Y is that first item on their initial list. If the informant responds with more items, you take it another step: "Try to remember other types of X like Y and tell me any new ones that you haven't already said." Do this until the informant says there are no more items like Y. Then you repeat the exercise for the second item on the informant's initial list; and the third; and so on (box 10.2).

You'd be surprised at how much you can learn from a humble set of free lists. Henley (1969) asked 21 students at Johns Hopkins University to name as many animals as they could in 10 minutes. She found an enormous variety of expertise when it comes to naming animals. In just this small group of informants (which didn't even represent the population of Johns Hopkins University, much less that of Baltimore or the United States), the lists ranged in length from 21 to 110, with a median of 55.

In fact, those 21 people named 423 different animals, and 175 were mentioned just once. The most popular animals for this group of informants were: dog, lion, cat, horse, and tiger, all of which were named by more than 90% of informants. Only 29 animals were listed by more than half the informants, but 90% of those were mammals. By contrast, among the 175 animals named only once, just 27% were mammals.

But there's more. Previous research had shown that the 12 most commonly talked about animals in American speech are: bear, cat, cow, deer, dog, goat, horse, lion, mouse, pig, rabbit, and sheep. There are $n(n-1)/2$, or 66 possible unique pairs of 12 animals (dog-cat, dog-deer, horse-lion, mouse-pig, etc.). Henley examined each informant's list of animals, and found the difference in the order of listing for each of the 66 pairs.

BOX 10.2

OTHER PROBES

Brewer tested three other kinds of probes for free lists: redundant questioning, nonspecific prompting, and alphabetic cueing. Here's the redundant question that Brewer and his colleagues asked a group of IV-drug users:

> Think of all the different kinds of drugs or substances people use to get high, feel good, or think and feel differently. These drugs are sometimes called recreational drugs or street drugs. Tell me the names of all the kinds of these drugs you can remember. Please keep trying to recall if you think there are more kinds of drugs you might be able to remember. (Brewer et al. 2002:347; and see Brewer and Garrett 2001)

In nonspecific prompting you ask people "What other kinds of X are there?" after they've responded to your original question. You keep asking this question until people say they can't think of any more Xs. And in alphabetic cueing, you ask informants "what kinds of X are there that begin with the letter A?" . . . "With the letter B?" And so on.

That is, if an informant mentioned goats 12th on her list, and bears 32nd, then the distance between goats and bears, for that informant, was $32 - 12 = 20$. Henley standardized these distances (that is, she divided each distance by the length of an informant's list and multiplied by 100) and calculated the average distance, over all the informants, for each of the 66 pairs of animals.

The lowest mean distance was between sheep and goats (1.8). If you named sheep, then the next thing you named was probably goats; and if you named goats, then next thing you named was probably sheep. Most speakers of English (and other Western languages, for that matter) have heard the expression: "That'll separate the sheep from the goats." This part of Western culture was originally a metaphor for distinguishing the righteous from the wicked and then became a metaphor for separating the strong from the weak. The first meaning was mentioned in the Old Testament (Ezekiel 34:17), and then again around 600 years later in the New Testament (Matthew 25:31–33).

Henley's respondents were neither shepherds nor students of Western scriptural lore, but they all knew that sheep and goats somehow "go together." Free lists tell you *what goes with what*, but you need to dig to understand *why*. Cats and dogs were only 2 units apart in Henley's free lists—no surprise there, right?—while cats and deer were 56 units apart. Deer, in fact, are related to all the other animals on the list by at least 40 units of distance, except for rabbits, which are only 20 units away from deer.

Robert Trotter (1981) asked 378 Mexican Americans to name the *remedios caseros*, or home remedies, they knew, and what illnesses each remedy was for. Informants listed a total of 510 remedies for treating 198 illnesses. However, the 25 most frequently mentioned remedies—about 5% of the 510—made up about 41% of all the cases; and the 70 most frequently mentioned illnesses—about 14%—made up 84% of the cases.

Trotter's free-list data reveal a lot about Mexican American perceptions of illness and

home cures. He was able to count which ailments were reported more frequently by men and which by women; which ailments were reported more frequently by older people and which by younger people; which by those born in Mexico and which by those born in the United States; and so on.

Informants who are very knowledgeable about the contents of a cultural domain usually provide longer lists than others. Some items will be mentioned over and over again, but eventually, if you keep asking people to list things, you get a lot of repeat items and all the new items are unique—that is, mentioned by only one informant. This happens pretty quickly (by the time you've interviewed 15 or 20 informants) with domains like names of ethnic groups, which are pretty well formed. With fuzzy domains, like "things that mothers do," you might still be eliciting new items after interviewing 30 or 40 people. Long lists don't necessarily mean that people know a lot about the things they name. In fact, in modern societies, people can often name a lot more things than they can recognize in the real world (see box 10.3).

BOX 10.3

LOOSE TALK

John Gatewood (1983) asked 40 adult Pennsylvanians to name all the trees they could think of. Then he asked them to check the trees on their list that they thought they could recognize in the wild. Thirty-seven of them listed "oak," 34 listed "pine," 33 listed "maple," and 31 listed "birch." I suspect that the list of trees and what people say they could recognize would look rather different in, say Wyoming or Mississippi. We could test that.

Thirty-one of the 34 who listed "pine" said they could recognize a pine. Twenty-seven people listed "orange," but only four people said they could recognize an orange tree without oranges hanging all over it. On average, the Pennsylvanians in Gatewood's sample said they could recognize half of the trees they listed, a phenomenon that Gatewood calls **loose talk**. He thinks that many Americans can name a lot more things than they can recognize in nature.

Does this loose talk phenomenon vary by gender? Suppose, Gatewood says, we ask Americans from a variety of subcultures and occupations to list other things besides trees. Would the 50% recognition rate hold? Gatewood and a group of students at Lehigh University asked 54 university students, half women and half men, to list all the musical instruments, fabrics, hand tools and trees they could think of. Then the informants were asked to check off the items in each of their lists that they thought they would recognize in a natural setting.

Gatewood chose musical instruments with the idea that there would be no gender difference in the number of items listed or recognized; he thought that women might name more kinds of fabrics than would men and that men would name more kinds of hand tools than would women. He chose the domain of trees to see if his earlier findings would replicate. All the hypotheses were supported (Gatewood 1984).

A. Kimball Romney and Roy D'Andrade asked 105 American high school students to "list all the names for kinds of relatives and family members you can think of in English" (1964:155). They were able to do a large number of analyses on these data. For example,

they studied the order and frequency of recall of certain terms, and the productiveness of modifiers, such as "step-," "half-," "-in-law," "grand-," "great," and so on. They assumed that the nearer to the beginning of a list a kin term occurs, the more salient it is for that particular informant. By taking the average position in all the lists for each kin term, they were able to derive a rank order list of kin terms, according to the variable's saliency.

They also assumed that more salient terms occur more frequently. So, for example, "mother" occurs in 93% of all lists and is the first term mentioned on most lists. At the other end of the spectrum is "grandson," which was only mentioned by 17% of the 105 informants, and was, on average, the 15th, or last term to be listed. They found that the terms "son" and "daughter" occur on only about 30% of the lists. But remember, these informants were all high school students, all of whom *were* sons and daughters, but none of whom *had* sons or daughters. It would be interesting to repeat Romney and D'Andrade's experiment on many different American populations. We could then test the saliency of English kin terms on the many subpopulations.

Finally, free listing can be used to find out where to concentrate effort in applied research, especially in rapid assessment. Researchers interested in high-risk sexual behavior, for example, use the free-list technique to understand domains like "ways to have sex" (Schensul et al. 1994) and "reasons to have sex" (Flores et al. 1998).

Monárrez-Espino et al. (2004) worked on a food aid program for at-risk Tarahumara infants in Mexico. A government agency had developed a basket of nutritional foods for distribution to Tarahumara mothers, but many of the foods (like canned sardines) were culturally unacceptable. Free listing of foods helped set things right.

In a project on which I consulted, interviewers asked people on the North Carolina coast how they viewed the possibility of offshore oil drilling. One of the questions was: "What are the things that make life good around here?" This question cropped up after some informal interviews in seven small, seaside towns. People kept saying "What a nice little town this is" and "What a shame it would be if things changed around here." Informants had no difficulty with the question, and after just 20 interviews, the researchers had a list of over 50 "things that make life good around here." The researchers chose the 20 items mentioned by at least 12 informants and explored the meaning of those items further (ICMR 1993).

The humble free list has many uses. Use it a lot (**Further Reading:** free lists).

THE TRUE-FALSE/YES-NO AND SENTENCE FRAME TECHNIQUES

Another common technique in cultural domain analysis is called the **sentence frame** or **frame elicitation** method. Linda Garro (1986) used the frame elicitation method to compare the knowledge of curers and noncurers in Pichátaro, Mexico. She used a list of 18 illness terms and 22 causes, based on prior research in Pichátaro (Young 1978). The frames were questions, like "can _____ come from _____?" Garro substituted names of illnesses in the first blank, and things like "anger," "cold," "overeating," and so on in the second blank. (ANTHROPAC has a routine for building questionnaires of this type.) This produced an 18 × 22 yes-no matrix for each of the informants. The matrices could then be added together and submitted to analysis by multidimensional scaling (see chapter 16).

James Boster and Jeffrey Johnson (1989) used the frame-substitution method in their study of how recreational fishermen in the United States categorize ocean fish. They asked 120 fishermen to consider 62 belief frames, scan down a list of 43 fish (tarpon, silver

perch, Spanish mackerel, etc.), and pick out the fish that fit each frame. Here are a few of the belief frames:

The meat from _____ is oily tasting.
It is hard to clean _____.
I prefer to catch _____.

That's 43 × 62 = 2,666 judgments by each of 120 informants, but informants were usually able to do the task in about half an hour (Johnson, personal communication). The 62 frames, by the way, came straight out of ethnographic interviews where informants were asked to list fish and to talk about the characteristics of those fish.

Gillian Sankoff (1971) studied land tenure and kinship among the Buang, a mountain people of northeastern New Guinea. The most important unit of social organization among the Buang is the *dgwa*, a kind of descent group, like a clan. Sankoff wanted to figure out the very complicated system by which men in the village of Mambump identified with various *dgwa* and with various named garden plots.

The Buang system was apparently too complex for bureaucrats to fathom, so, to save administrators a lot of trouble, the men of Mambump had years earlier devised a simplified system that they presented to outsiders. Instead of claiming that they had ties with one or more of five different *dgwa*, they each decided which of the two largest *dgwa* they would belong to, and that was as much as the New Guinea administration knew.

To unravel the complex system of land tenure and descent, Sankoff made a list of all 47 men in the village and all 140 yam plots that they had used over the recent past. Sankoff asked each man to go through the list of men and identify which *dgwa* each man belonged to. If a man belonged to more than one, then Sankoff got that information, too. She also asked her informants to identify which *dgwa* each of the 140 garden plots belonged to.

As you might imagine, there was considerable variability in the data. Only a few men were uniformly placed into one of the five *dgwa* by their peers. But by analyzing the matrices of *dgwa* membership and land use, Sankoff was able to determine the core members and peripheral members of the various *dgwa*.

She was also able to ask important questions about intracultural variability. She looked at the variation in cognitive models among the Buang for how land use and membership in descent groups were related. Sankoff's analysis was an important milestone in our understanding of the measurable differences between individual culture versus shared culture. It supported Goodenough's notion (1965) that cognitive models are based on shared assumptions, but that ultimately they are best construed as properties of individuals.

Techniques like true-false and yes-no tests that generate nominal data are easy to construct, especially with ANTHROPAC, and can be administered to a large number of informants. Frame elicitation in general, however, can be boring, both to the informant and to the researcher alike. Imagine, for example, a list of 25 animals (mice, dogs, antelopes . . .), and 25 attributes (ferocious, edible, nocturnal . . .).

The structured interview that results from such a test involves a total of 625 (25 × 25) questions to which an informant must respond—questions like "Is an antelope edible?" "Is a dog nocturnal?" "Is a mouse ferocious?" People can get pretty exasperated with this kind of foolishness, so be careful to choose domains, items, and attributes that make sense to people when you do frame elicitations and true-false tests (**Further Reading:** sentence frames).

TRIAD TESTS

In a **triad test**, you show people three things and tell them to "Choose the one that doesn't fit" or "Choose the two that seem to go together best," or "Choose the two that are the same." The "things" can be photographs, dried plants, or 3 × 5 cards with names of people on them. (Respondents often ask "What do you mean by things being 'the same' or 'fitting together'?" Tell them that you are interested in what *they* think that means.) By doing this for all triples from a list of things or concepts, you can explore differences in cognition among individuals, and among cultures and subcultures.

Suppose you ask speakers of English to "choose the item that is least like the other two" in each of the following triads:

DOLPHIN	MOOSE	WHALE
SHARK	DOLPHIN	MOOSE

All three items in the first triad are mammals, but two of them are sea mammals. Some native speakers of English will choose "dolphin" as the odd item out because "whales and moose are both big mammals and the dolphin is smaller." In my experience, though, most people will choose "moose" as the most different because "whales and dolphins are both sea animals." In the second triad, many of the same people who chose "moose" in the first triad will choose "shark" because moose and dolphins are both mammals and sharks are not.

But some people who chose "moose" in triad 1 will choose "moose" again because sharks and dolphins are sea creatures, while moose are not. Giving people a judiciously chosen set of triad stimuli can help you understand interindividual similarities and differences in how people think about the items in a cultural domain (box 10.4).

Lieberman and Dressler (1977) used triad tests to examine intracultural variation in ethnomedical beliefs on the Caribbean island of St. Lucia. They wanted to know if cognition of disease terms varied with bilingual proficiency. They used 52 bilingual English-Patois speakers, and 10 monolingual Patois speakers. From ethnographic interviewing and cross-checking against various informants, they isolated nine disease terms that were important to St. Lucians.

Here's the formula for finding the number of triads in a list of n items:

$$\frac{n(n-1)(n-2)}{6} \qquad \textbf{Formula 10.1}$$

In this case, $n = 9$ (the number of disease terms), so there are 84 possible triads.

Lieberman and Dressler gave each of the 52 bilingual informants two triad tests, a week apart: one in Patois and one in English. (Naturally, they randomized the order of the items within each triad and randomized the order of presentation of the triads to informants.) They also measured how bilingual their informants were, using a standard test. The 10 monolingual Patois informants were simply given the triad test.

The researchers counted the number of times that each possible pair of terms was chosen as most alike among the 84 triads. (There are $n \times n-1/2$ pairs or $9 \times 8/2 = 36$ pairs). They divided the total by seven (the maximum number of times that any pair appears in the 84 triads). This produced a similarity coefficient, varying between 0.0 and 1.0, for each possible pair of disease terms. The larger the coefficient for a pair of terms, the closer in meaning the two terms are. The researchers were then able to analyze these data among English-dominant, Patois-dominant, and monolingual Patois speakers. (I'll show you how to analyze triad test data in chapter 16.)

BOX 10.4

TRIAD TESTS AND COGNITIVE SCIENCE

The triads test was developed in psychology (see Kelly 1955; Torgerson 1958) and has long been used in studies of cognition. Romney and D'Andrade (1964) presented people with triads of American kinship terms and asked them to choose the term that was most *dissimilar* in each triad. For example, when they presented informants with the triad "father, son, nephew," 67% selected "nephew" as the most different of the three items. Twenty-two percent chose "father" and only 2% chose "son." Romney and D'Andrade asked people to explain *why* they'd selected each item on a triad. For the triad "grandson, brother, father," for example, one informant said that a "grandson is most different because he is moved down further" (p. 161). There's a lot of cultural wisdom in that statement.

By studying which pairs of kinship terms their informants chose most often as being as similar, Romney and D'Andrade were able to isolate some of the salient components of the American kinship system (components such as male vs. female, ascending vs. descending generation, etc.). They were able to do this, at least, for the group of informants they used. Repeating their tests on other populations of Americans, or on the same population over time, would yield interesting comparisons.

It turned out that when Patois- and English-dominant informants took the triad test in English, their cognitive models of similarities among diseases was similar. When Patois-dominant speakers took the Patois-language triad test, however, their cognitive model was similar to that of monolingual Patois informants.

This is a very interesting finding. It means that Patois-dominant bilinguals manage to hold on to two distinct psychological models about diseases and switch back and forth between them, depending on what language they are speaking. By contrast, the English-dominant group displayed a similar cognitive model of disease terms, irrespective of the language in which they are tested.

The Balanced Incomplete Block Design for Triad Tests

Typically, the terms that go into a triad test are generated by a free list, and typically the list is much too long for a triad test. As you can see from formula 10.1, with just 9 terms, there are 84 stimuli in a triad test containing nine items. But with 15 items, just 6 more, the number of decisions an informant has to make jumps to 455. At 20 items, it's a mind-numbing 1,140.

Free lists of illnesses, ways to prevent pregnancy, advantages of breast-feeding, places to go on vacation, and so on easily produce 60 items or more. Even a selected, abbreviated list may be 20 items.

This led Michael Burton and Sara Nerlove (1976) to develop the **balanced incomplete block design**, or BIB, for the triad test. BIBs take advantage of the fact that there is a lot of redundancy in a triad test. Suppose you have just four items, 1, 2, 3, 4 and you ask informants to tell you something about *pairs* of these items (e.g., if the items were vegeta-

bles, you might ask "Which of these two is less expensive?" or "Which of these two is more nutritious?" or "Which of these two is easier to cook?"). There are exactly six pairs of four items (1–2, 1–3, 1–4, 2–3, 2–4, 3–4), and the informant sees each pair just once.

But suppose that instead of pairs you show the informant triads and ask which two out of each triple are most similar. There are just four triads in four items (1–2–3, 1–2–4, 2–3–4, 1–3–4), but each item appears $(n-1)(n-2)/2$ times, and each pair appears $n-2$ times. For four items, there are $n(n-1)/2 = 6$ pairs; each pair appears twice in four triads, and each item on the list appears three times.

It is all this redundancy that reduces the number of triads needed in a triads test. In a complete set of 84 triads for 9 items, each pair of items appears $n-2$, or seven times. If you have each pair appear just once (called a **lambda 1 design**), instead of seven times, then, instead of 84 triads, only 12 are needed. If you have each pair to appear twice (a lambda 2 design), then 24 triads are needed. For analysis, a lambda 2 design is much better than a lambda 1. Table 10.1 shows the lambda 2 design for 9 items and 10 items.

Table 10.1 Balanced Incomplete Block Designs for Triad Tests Involving 9 and 10 Items

For 9 items, 24 triads are needed, as follows:		*For 10 items, 30 triads are needed, as follows:*	
Items		Items	
1, 5, 9	1, 2, 3	1, 2, 3	6, 8, 9
2, 3, 8	4, 5, 6	2, 5, 8	7, 10, 3
4, 6, 7	7, 8, 9	3, 7, 4	8, 1, 10
2, 6, 9	1, 4, 7	4, 1, 6	9, 5, 2
1, 3, 4	2, 5, 9	5, 8, 7	10, 6, 7
5, 7, 8	3, 6, 8	6, 4, 9	1, 3, 5
3, 7, 9	1, 6, 9	7, 9, 1	2, 7, 6
2, 4, 5	2, 4, 8	8, 10, 2	3, 8, 9
1, 6, 8	3, 5, 7	9, 3, 10	4, 2, 10
4, 8, 9	1, 5, 8	10, 6, 5	5, 6, 3
3, 5, 6	2, 6, 8	1, 2, 4	6, 1, 8
1, 2, 7	3, 4, 9	2, 3, 6	7, 9, 2
		2, 4, 8	8, 4, 7
		4, 9, 5	9, 10, 1
		5, 7, 1	10, 5, 4

SOURCE: Reprinted from *Social Science Research*, Vol. 5, M. L. Burton and S. B. Nerlove, "Balanced Design for Triad Tests," p. 5, © 1976. Reprinted by permission of Academic Press.

For 10 items, a lambda 2 design requires 30 triads; for 13 items, it requires 52 triads; for 15 items, 70 triads; for 19 items, 114 triads; and for 25 items, 200 triads. Unfortunately, there is no easy formula for choosing *which* triads in a large set to select for a BIB. Fortunately, Burton and Nerlove (1976) worked out various lambda BIB designs for up to 21 items and Borgatti has incorporated BIB designs into ANTHROPAC (1992a). You simply tell ANTHROPAC the list of items you have, select a design, and tell it the number of informants you want to interview. ANTHROPAC then prints out a randomized triad test, one for each informant. (Randomizing the order in which the triads appear to informants eliminates "order-effects"—possible biases that come from responding to a list of stimuli in a particular order.)

Boster et al. (1987) used a triad test in their study of the social network of an office. There were 16 employees, so there were 16 "items" in the cultural domain ("the list of all the people who work here" is a perfectly good domain). A lambda 2 test with 16 items has 80 distinct triads. Informants were asked to "judge which of three actors was the most different from the other two."

Triad tests are easy to create with ANTHROPAC, easy to administer, and easy to score, but they can only be used when you have relatively few items in a cultural domain. In literate societies, most informants can respond to 200 triads in less than half an hour, but it can be a really boring exercise, and boring your informants is a really bad idea. I find that informants can easily handle lambda 2 triad tests with up to 15 items and 70 triads. But I also find that people generally prefer—even like—to do pile sorts (**Further Reading:** triad tasks).

FREE PILE SORTS

In 1966, John Brim put the names of 58 American English role terms (mother, gangster, stockbroker, etc.) on slips of paper. He asked 108 high school students in San Mateo, California, to spread the slips out on their desks and to "put the terms together which you feel belong together" (Burton and Romney 1975:400). This simple, compelling method for collecting data about what-goes-with-what was introduced to anthropology by Michael Burton, who analyzed Brim's data using multidimensional scaling and hierarchical clustering. These powerful tools were brand new at the time and are used today across the social sciences (Burton 1968, 1972). (We'll get back to MDS and clustering in chapter 16 on how to analyze data in cultural domains.)

I've used **free pile sorts** to study the social structure of institutions such as prisons, ships at sea, and bureaucracies, and also to map the cognitively defined social organization of small communities. I simply hand people a deck of cards, each of which contains the name of one of the people in the institution, and ask informants to sort the cards into piles, according to their own criteria. The results tell me how people in the various components of an organization (managers, production workers, advertising people; or guards, counselors, prisoners; or seamen, deck officers, engine room personnel; or men and women in a small Greek village) think about the social structure of the group. Instead of what goes with what, I learn who goes with whom. Then I ask informants to explain *why* people appear in the same pile. This produces a wealth of information about the cognitively defined social structure of a group.

Administering a Pile Sort

Informants often ask two questions when asked to do a pile sort: (1) "What do you mean by 'belong together'?" and (2) "Can I put something in more than one pile?" The answer to the first question is "There are no right or wrong answers. We want to learn what *you* think about these things."

The easy answer to the second question is "No," because there is one card per item and a card can only be in one pile at a time. This answer cuts off a lot of information, however, because people can think of items in a cultural domain along several dimensions at once. For example, in a pile sort of consumer electronics, someone might want to put DVD recorders in one pile with TVs (for the obvious association) and in another pile with camcorders (for another obvious association), but might not want to put camcorders and TVs in the same pile. One way to handle this problem is to have duplicate cards that you can give to people when they want to put an item into more than one pile, but be warned that this can complicate analysis of the data. An alternative is to ask the informant to do **multiple free pile sorts** of the same set of items (box 10.5).

The P-3 Game

In a series of papers, John Roberts and his coworkers used pile sorts and rating tasks to study how people perceive various kinds of behaviors in games (see, for example,

BOX 10.5

PILE SORTS WITH OBJECTS

Pile sorts don't have to be done with cards. James Boster (1987) studied the structure of the domain of birds among the Aguaruna Jívaro of Peru. He paid people to bring him specimens of birds and he had the birds stuffed. He built a huge table out in the open, laid the birds on the table, and asked the Aguaruna to sort the birds into groups.

Carl Kendall led a team project in El Progreso, Honduras, to study beliefs about dengue fever (Kendall et al. 1990). Part of their study involved a pile sort of the nine most common flying insects in the region. They mounted specimens of the insects in little boxes and asked people to group the insects in terms of "those that are similar." Some fieldworkers have used photographs of objects as stimuli for a pile sort.

Borgatti (1999:133), however, points out that physical stimuli, like images or objects, make people focus on form rather than function. In fact, when asked to sort drawings of fish, fishermen in North Carolina sorted on shape—the long thin ones, the ones with a big dorsal fin, the small roundish ones (Boster and Johnson 1989). "In contrast," says Borgatti (1999:133), "sorting *names* of fish allows hidden attributes to affect the sorting"—things like taste or how much of a struggle fish put up. "If you are after shared cultural beliefs," says Borgatti, "I recommend keeping the stimulus as abstract as possible" (1992b:6).

Roberts and Chick 1979; Roberts and Nattrass 1980). One "game," studied by Roberts et al. (1980), is pretty serious: searching for foreign submarines in a P-3 airplane. The P-3 is a four-engine, turboprop, low-wing aircraft that can stay in the air for a long time and cover large patches of ocean. It is also used for search-and-rescue missions. Making errors in flying the P-3 can result in career damage and embarrassment, at least, and injury or death, at worst.

Through extensive, unstructured interviews with Navy P-3 pilots, Roberts et al. isolated 60 named flying errors. (This is the equivalent of extracting a free list from your interviews.) Here are a few of the errors: flying into a known thunderstorm area; taking off with the trim tabs set improperly; allowing the prop wash to cause damage to other aircraft; inducing an autofeather by rapid movement of power level controls. Roberts et al. asked 52 pilots to do a free pile sort of the 60 errors and to rate each error on a 7-point scale of "seriousness."

They also asked the pilots to rank a subset of 13 errors on four criteria: (1) how much each error would "rattle" a pilot; (2) how badly each error would damage a pilot's career; (3) how embarrassing each error would be to commit; and (4) how much "fun" it would be to commit each error. Flying into a thunderstorm on purpose, for example, could be very damaging to a pilot's career, and extremely embarrassing if he had to abort the mission and turn back in the middle (when Roberts et al. did their research in the 1970s, all P-3 pilots were men). But if the mission was successful, then taking the risk of committing a very dangerous error would be a lot of fun for pilots who are, as Roberts called them, "high self-testers" (personal communication).

Inexperienced pilots rated "inducing an autofeather" as more serious than did highly experienced pilots. Inducing an autofeather is more embarrassing than it is dangerous and it's the sort of error that experienced pilots just don't make. On the other hand, as the number of air hours increased, so did pilots' view of the seriousness of "failure to use all available navigational aids to determine position." Roberts et al. suggested that inexperienced pilots might not have had enough training to assess the seriousness of this error correctly (**Further Reading:** pile sorts).

The Lumper-Splitter Problem

In the free pile sort method, people are told that they can make as many piles as they like, so long as they don't make a separate pile for each item or lump all the items into one pile. Like the triad test, the free pile sort presents people with a common set of stimuli, but there's a crucial difference: With free pile sorts, people can group the items together as they see fit. The result is that some people will make many piles, others will make few, and this causes the **lumper-splitter** problem (Weller and Romney 1988:22).

In a pile sort of animals, for example, some informants will put all the following together: giraffe, elephant, rhinoceros, zebra, wildebeest. They'll explain that these are the "African animals." Others will put giraffe, elephant, and rhino in one pile, and the zebra and wildebeest in another, explaining that one is the "large African animal" pile and the other is the "medium-sized African animal pile."

Although they can't put *every* item in its own pile, lots of people put *some* items in singleton piles, explaining that each item is unique and doesn't go with the others. It's fine to ask informants why they made each pile of items, but wait until they finish the sorting task so you don't interfere with their concentration. And don't hover over informants. Find an excuse to walk away for a couple of minutes after they get the hang of it.

Because triad tests present each respondent with exactly the same stimuli, you can compare the data across individuals. Free pile sorts tell you what the structure of the data looks like for a group of people—sort of group cognition—but you can't compare the data from individuals. On the other hand, with pile sorts, you can have as many as 50 or 60 items. All methods have their advantages and disadvantages.

RANKINGS AND PAIRED COMPARISONS

Rank ordering produces interval-level data, while ratings ("on a scale of 1-to-5, how much do you like . . . ?") produce ordinal-level data. Not all behaviors or concepts are easy to rank and there are lots of times when ratings are the best you can do, but when you can get rank-ordered data you shouldn't pass up the opportunity. Eugene Hammel (1962) asked people in a Peruvian village to rank order the people they knew in terms of prestige. By comparing the lists from different informants, Hammel was able to determine that the men he tested all had a similar view of the social hierarchy. Occupations can easily be rank ordered on the basis of prestige, or lucrativeness.

Or even accessibility. The instructions to respondents would be "Here is a list of occupations. Please rank them in order, from most likely to least likely that your daughter will have this occupation." Then ask respondents to do the same thing for their sons. (Be sure to assign people randomly to doing the task for sons or daughters first.) Then compare the average ranking of accessibility against some independent variables and test for intracultural differences among ethnic groups, genders, age groups, and income groups.

Weller and Dungy (1986) studied breast-feeding among Hispanic and Anglo women in southern California. They asked 55 informants for a free list of positive and negative

aspects of breast- and bottle-feeding. Then they selected the 20 most frequently mentioned items in this domain and converted the items to neutral, similarly worded statements. A few examples: "A way that doesn't tie you down, so you are free to do more things"; "A way that your baby feels full and satisfied"; "A way that allows you to feel closer to your baby."

Next, Weller and Dungy asked 195 women to rank the 20 statements. The women were asked which statement was most important to them in selecting a method of feeding their baby, which was the next most important to them, and so on. In the analysis, Weller and Dungy were able to relate the average rank order for Hispanics and for Anglos to independent variables like age and education.

Paired Comparisons

The method of **paired comparisons** is an alternative way to get rank orderings of a list of items in a domain. For any set of things, there are $n(n-1)/2$ pairs of those things. Suppose you have a list of five colors: red, green, yellow, blue, and brown. Figure 10.1 shows the paired comparison test to find out an informant's rank-ordered preference for these five colors. In this case, the question would be: "Look at each pair of colors and, for each pair, tell me which one you like more."

In each of the following pairs of colors, please circle the one you like best:

RED	GREEN
RED	YELLOW
RED	BLUE
RED	BROWN
GREEN	YELLOW
GREEN	BLUE
GREEN	BROWN
YELLOW	BLUE
YELLOW	BROWN
BLUE	BROWN

FIGURE 10.1.
A paired comparison test for rank-ordered data.

You might say: "Here are two animals. Which one is the more _____?," where the blank is filled in by "vicious," or "wild," or "smarter," or some other descriptor.

You could ask informants to choose "the food in this pair that is better for you," or "the crime in this pair that you're most afraid of."

I've presented the pairs in figure 10.1 in such a way that you can easily see how the 10 of them exhausts the possibilities for five items. When you present a paired comparison test to an informant, be sure to scramble the order of the pairs to guard against **order effects**—that is, where something about the order of the items in a list influences the choices that informants make.

To find the rank order of the list for each informant, you simply count up how many times each item in a list "wins"—that is, how many times it was circled. If you are studying illnesses and cancer is on the list, and if the question is "which of these pairs of illnesses is more life threatening," you expect to find it circled each time it is paired with another illness—except, perhaps, when it is paired with AIDS. Because this is so predictable, it's not very interesting. It gets really interesting when you have illnesses like diabetes and high blood pressure in your list and you compare the average rank ordering among various ethnic groups.

The paired comparison technique has a lot going for it. People make one judgment at a time, so it's much easier on them than asking them to rank order a list of items by staring at all the items at once. Also, you can use paired comparisons with nonliterate informants by reading the list of pairs to them, one at a time, and recording their answers.

Like triad tests, paired comparisons can only be used with a relatively short list of items in a domain, unless you apply balanced incomplete block designs. With 20 items in a paired comparison task, for example, informants have to make 190 judgments (**Further Reading:** rankings and paired comparisons).

There is one more method for studying the attributes of things: rating scales. This one is so important, it deserves a chapter of its own . . . next.

FURTHER READING

Kinship studies: Alexander (1976); Dousset (2008); Houseman and White (1998); Kronenfeld (2004); Leach (1945); Lehman (1992); Read (2001).

Methods of data collection for domain analysis: de Munck and Sobo (1998); Handwerker (2001); J. C. Johnson and Weller (2002); Weller and Romney (1988).

Free lists: Ross and Medin (2005); Ryan et al. (2000); K. D. Smith et al. (2007); Thompson and Juan (2006); Verma et al. (2001).

Sentence frames: D'Andrade et al. (1972); Frake (1964); Hruschka et al. (2008); Metzger and Williams (1966).

Triad tasks: Durrenberger and Erem (2005); Furlow (2003); Nyamongo (2002); Ross et al. (2005).

Pile sorts: Collins (2006); Longfield (2004); Roberts et al. (1986).

Rankings and paired comparisons: Chavez et al. (1995); Durrenberger (2003); Erickson (1997); Kozak et al. (2008); Thurstone (1927).

16

Cognitive Anthropology I: Analyzing Cultural Domains

In chapter 10, I introduced you to five methods for collecting systematic data about cultural domains: free lists, sentence frames, triad tests, pile sorts, and paired comparisons. In this chapter, I'll show you how to analyze these kinds of data. In the next chapter, I'll show you three more methods that are grounded in cognitive anthropology: componential analysis, folk taxonomies, and ethnographic decision modeling. We begin with free lists (box 16.1).

ANALYZING FREE LISTS

Gery Ryan and I asked 34 people: "Please write down the names all the fruits you can think of" (Bernard and Ryan 2010:167). Because free list data are texts, they have to be cleaned up before you can analyze them. Only 10 people listed grapes, but another 22 (for a total of 32 out of 34 people) listed grape (in the singular). Before counting up the frequency for each item in the free lists, we had to combine all mentions of grapes and grape. It doesn't matter whether you change grapes into grape or vice versa, so long as you make all the required changes.

It takes some work to clean up the spelling in free lists. In our data, three people listed bananna (wrong spelling), and 27 people listed banana (right spelling); three people listed avacado (wrong), one listed avocato (wrong), and six people listed avocado (right). Cantaloupe was hopeless, as was pomegranate. We got eight cantaloupe (the preferred spelling in the dictionary), six cantelope, two cantelopes, and three canteloupe. We got 17 listings for guava and one for guayaba, which happens to be the Spanish term for guava. We got 10 listings for passion fruit and one for passion-fruit, with a hyphen (when computers list word frequencies, they see those two listings as different).

Once the data were cleaned, we plotted how often each fruit was mentioned. The result is the scree plot in figure 16.1. ("Scree" refers to the rocks that pile up at the base of a cliff and the telltale L-shape of the scree.)

The shape of the curve in figure 16.1 is typical for a well-defined domain, like fruits: The 34 informants named a total of 147 different fruits, but 88 of those fruits were named by just one person (prickly pear and quince, for example) and almost everyone named a few items (apple and orange, for example). Compare that to the results for lists of "things that mothers do." For this domain, our 34 informants named 554 items, of which 515 were named by just one person and only a handful (love, clean, cook) were named by five or more people.

The difference is that fruits (and animals, and names of racial/ethnic groups, and emo-

BOX 16.1

ABOUT SKILLS IN DATA ANALYSIS

We could begin our tour of data analysis anywhere, with any kind of data, because all data analysis, whether we're talking about the qualitative or quantitative kind, requires the same set of big skills. Here's a list of those big skills: (1) logical reasoning; (2) ways to test the results of logical reasoning. That's it. That's the whole list of big skills. All the other skills (and there are hundreds of them) are in one of the two big-skill categories.

The ability to engage in constant comparison, for example, is a really nice skill to develop. It's part of the logical-reasoning set of tools and it means being able to hold some ideas in your head so you can compare them to new ones that emerge as you analyze data. For qualitative data, this means holding on to themes in some narratives and comparing them to new themes that pop into your head as you read the narratives a second and a third and a fourth time. For quantitative data, constant comparison means looking for patterns in many small findings as you run, say, correlations or chi-squared tests to examine hypotheses about what goes with what.

The point is, any skills you learn about how to analyze any data—qualitative or quantitative—are going to serve you well as you work with more and different kinds of data. In fact, having lots of skills in analyzing data will make you fearless about collecting different kinds of data. Once you master multidimensional scaling and cluster analysis, for example—two methods of analysis we'll cover in this chapter—you'll be able to analyze network data and once you can do that, you can do semantic network analysis, which will dramatically extend your capacity to analyze text, and . . . you get the idea (appendix E).

tions) are very well defined, but things that mothers do (and things that people might do on a weekend, and things that you can do to stay healthy) are much less well-defined cultural domains. Many of the most interesting domains are things that people don't have easy lists for.

Measuring the Salience of Free-List Items

Gery Ryan and I asked 42 American adolescents (20 boys and 22 girls) to list things they were worried about concerning their health (Bernard and Ryan 2010:170). Table 16.1 shows the results.

Over three-quarters of the informants (76.2%) mentioned sexually transmitted diseases (STDs), and over a third (35.7%) specifically mentioned HIV/AIDS. This is what we expect from adolescents but, surprisingly, nearly half (45.2%) of the informants (all under age 20) were worried about cancer—a worry usually associated with older people.

When we explored this, we found that just 6 of the 20 boys in our sample (30%) had mentioned cancer, compared to 13 of the 22 girls (59%). And when the boys mentioned cancer at all, they ranked it fifth, on average, of the illnesses they were worried about, compared to second for the girls. The girls, it turned out, were very worried about breast cancer, but when the data from both genders were combined, this wasn't noticeable.

The frequency of items in a set of free lists is one indicator of the importance—or

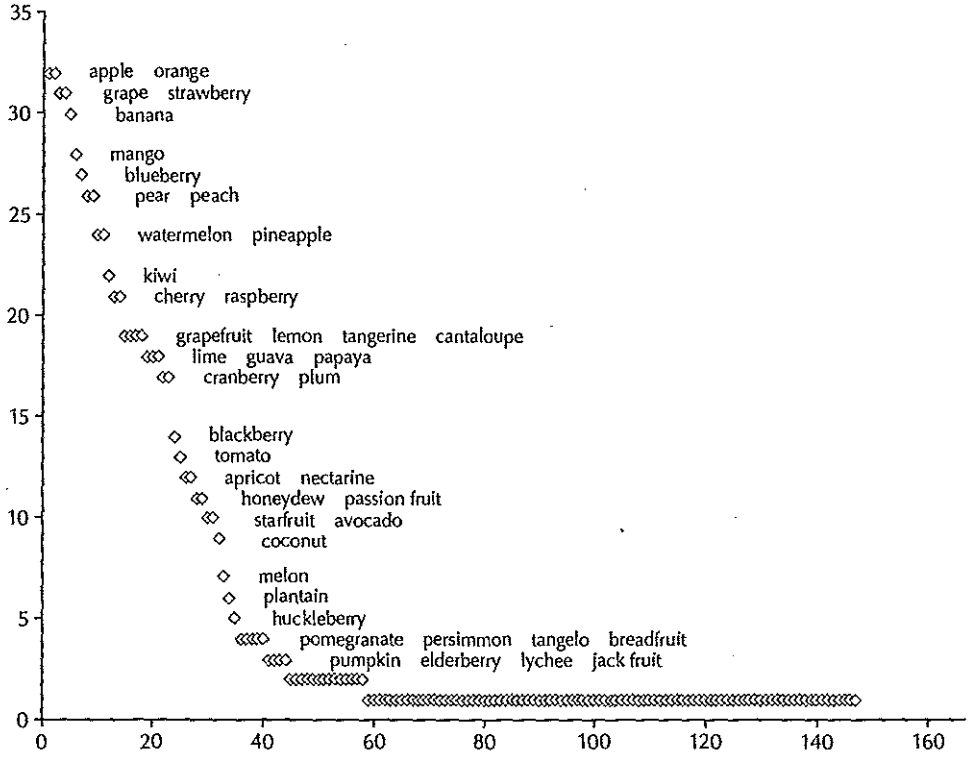

FIGURE 16.1.

Scree plot of free list of 143 fruits from 34 informants.

SOURCE: H. R. Bernard and G. W. Ryan, *Analyzing Qualitative Data: Systematic Approaches.* Los Angeles: Sage Publications. 2010. p. 169. Used by permission.

Table 16.1 Free List Results from 42 Adolescents about Their Health Concerns

	Total Sample (N = 42)		*Girls (n = 22)*			*Boys (n = 20)*			*Difference*
	Freq.	*%*	*Rank*	*Freq.*	*%*	*Rank*	*Freq.*	*%*	*women% – men%*
Cold/Flu	12	28.6	3	10	45.5	9	2	10.0	35.5
Cancer	19	45.2	2	13	59.1	5	6	30.0	29.1
Eating Disorders	10	23.8	4	8	36.4	9	2	10.0	26.4
HIV/AIDS	15	35.7	3	10	45.5	6	5	25.0	20.5
Mono	10	23.8	5	7	31.8	8	3	15.0	16.8
Stress	8	19.0	7	5	22.7	8	3	15.0	7.7
Weight-Obesity	10	23.8	6	6	27.3	7	4	20.0	7.3
Skin-related	13	31.0	5	7	31.8	5	6	30.0	1.8
Hygiene	8	19.0	8	4	18.2	7	4	20.0	−1.8
Disease	8	19.0	8	4	18.2	7	4	20.0	−1.8
Eating Right	7	16.7	9	3	13.6	7	4	20.0	−6.4
STDs	32	76.2	1	16	72.7	1	16	80.0	−7.3
Fitness	12	28.6	7	5	22.7	4	7	35.0	−12.3
Drug Abuse	9	21.4	9	3	13.6	5	6	30.0	−16.4
Alcohol-related	16	38.1	6	6	27.3	2	10	50.0	−22.7
Smoking-related	11	26.2	9	3	13.6	3	8	40.0	−26.4

SOURCE: H. R. Bernard and G. W. Ryan, *Analyzing Qualitative Data: Systematic Approaches.* Los Angeles: Sage Publications. 2010. p. 171. Used by permission.

salience—of those items to informants. Another indicator is how early, on average, an item gets mentioned. If you ask native speakers of American English to list animals, you'll find that (1) cat and dog are mentioned a lot, and (2) they are mentioned early (Henley 1969). In fact, those two animals are typically the first two animals that get mentioned.

Charismatic megafauna—elephants, whales, lions, and so on—also get mentioned a lot, but usually after the common household animals get named. Thus, in addition to frequency, we can measure the average rank that each item appears in a set of lists. Free listing, however, produces lists of varying length. It's one thing to name elephants fifth in a list of 30 animals and quite another to name elephants fifth in a list of 10 animals, and several methods for taking these factors into account are available.

Smith's S (J. J. Smith and Borgatti 1997) takes into account both the frequency of an item and how early in each list is mentioned and is a popular measure of item cognitive salience. It is also "highly correlated with simple frequency" (Borgatti 1999:149) and so, for most analyses, simple frequency counts of free list data are all that's needed. ANTHROPAC software (Borgatti 1992a) makes short work of free lists (see appendix E) (**Further Reading**: measuring salience. See **Further Reading** on free lists in chapter 10.)

Selecting Items from a Free List for Further Study

Researchers use scree plots to choose a set of items to study in more depth. For example, by counting the dots in figure 16.1, we see that (1) 14 fruits were mentioned by 20 or more of our 34 informants, and (2) 58 items were mentioned by at least two of our informants. All the other fruits were mentioned just once.

How many items should we choose from these data as representing the contents of the domain? There is no formal rule here, but a good general rule is to select items that are mentioned by at least 10% of your informants. If you have 40 informants, then choose items that were mentioned by at least four of them. If this still produces too many items, then move up to 15% of informants or more.

There is nothing forcing you to take every item that's mentioned a lot, especially if you already know something about the domain you're studying. If you want to study, say, 40 items in depth, you can choose some that are mentioned frequently and others that are mentioned less frequently—or even by no one at all.

An item mentioned once is usually not a good candidate to include for further work on the structure of the domain. The whole idea of a cultural domain, as contrasted with an individual cognitive domain is that the content is shared (Borgatti 1999). On the other hand, we often want to know where a particular item fits within a cultural domain.

Once we have identified the items in a cultural domain, the next step is to examine how the items are related to each other. To do this, we ask informants to make similarity judgments—to tell us what goes with what. Pile sorts are an effective method for collecting these judgments.

ANALYZING PILE SORT DATA

Begin a pile sort task by writing the name of each item on a single card (index cards work nicely). Label the back of each card with the number from 1 to n (where n is the total number of items in the domain). Spread the cards out randomly on a large table with the item-side up and the number-side down. (Be sure to shuffle the deck between informants.) Ask informants to sort the cards into piles according to which items they think belong together.

Figure 16.2 shows the pile sort data for one male informant who sorted the names of

1. Apple
2. Orange
3. Papaya
4. Mango
5. Peach
6. Blueberry
7. Watermelon
8. Pineapple
9. Pear
10. Strawberry
11. Lemon
12. Cantaloupe
13. Grapefruit
14. Plum
15. Banana
16. Avocado
17. Fig
18. Cherry

One Person's Sorting of 18 Fruits:
Pile #1: 2, 11, 13
Pile #2: 1, 5, 9, 14, 17, 18
Pile #3: 3, 4, 8, 15, 16
Pile #4: 6, 10
Pile #5: 7, 12

FIGURE 16.2.
Pile sort data from one person for 18 fruits.
SOURCE: H. R. Bernard and G. W. Ryan, *Analyzing Qualitative Data: Systematic Approaches*. Los Angeles: Sage Publications. 2010. p. 173. Used by permission.

18 fruits. It also shows the format for recording the pile sort data. Pile #1 contained items 2, 11, and 13. In other words, this informant put the cards for orange, lemon, and grapefruit into one pile.

Table 16.2 shows the data from figure 16.2 in the form of a similarity matrix, similar to the one you saw in chapter 15.

When the informant put items 2, 11, and 13 (orange, lemon, grapefruit) into a pile, he did so because he thought the items were similar. To indicate this, there is a 1 in the matrix where items 2 and 11 intersect; another 1 in the cell where items 2 and 13 intersect; and another 1 in the cell where 11 and 13 intersect.

Similarly for Pile #2: There is a 1 in the 1-5 cell, the 1-9 cell, the 1-14 cell, and so on. There are 0s in all the cells that represent no similarity of a pair of items (for this informant) and 1s down the diagonal (since items are similar to themselves). Notice that if 11 is similar to 13, then 13 is similar to 11, so this particular matrix is also symmetric. In a

Table 16.2 Similarity Matrix from One Person's Pile Sorting of the 18 Fruits

		1 AP	*2 OR*	*3 PA*	*4 MA*	*5 PE*	*6 BL*	*7 WA*	*8 PI*	*9 PE*	*10 ST*	*11 LE*	*12 CA*	*13 GR*	*14 PL*	*15 BA*	*16 AV*	*17 FI*	*18 CH*
1	APPLE	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
2	ORANGE	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
3	PAPAYA	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
4	MANGO	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
5	PEACH	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
6	BLUEBERRY	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
7	WATERMELON	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
8	PINEAPPLE	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
9	PEAR	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
10	STRAWBERRY	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
11	LEMON	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
12	CANTALOUPE	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
13	GRAPEFRUIT	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
14	PLUM	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
15	BANANA	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
16	AVOCADO	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
17	FIG	1	0	0	0	1	—	0	0	0	1	0	0	0	0	1	0	0	1	1
18	CHERRY	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1

symmetric matrix, the bottom and top halves (above and below the diagonal of 1s) are identical (box 16.2).

BOX 16.2

GETTING PILE-SORT DATA INTO A COMPUTER

ANTHROPAC is a DOS program, but it's still the easiest way I know to import pile sort data into a computer for analysis. Once you have pile sort data into a computer, you can use any major statistical package to analyze the matrices. I use UCINET, a windows program, because it shares files with ANTHROPAC. Any data you import with ANTHROPAC is available to UCINET for analysis. UCINET also can export data as an Excel file, so you can use the data in your favorite statistics program.

ANALYZING PILE SORT DATA: MDS

If you examine it carefully, you'll see that, despite the 1s and 0s, there is not a shred of math in table 16.2. It contains nothing more than the information in the bottom half of figure 16.2, displayed as 1s and 0s, and there is nothing numerical about those 1s and 0s. They simply stand for oranges and papayas and so on. But by substituting 1s and 0s for the names of the items, we can use software to look for patterns in the informant's pile sort data.

Figure 16.3 is a multidimensional scaling, or MDS, of these data. MDS is one of several visualizations methods now widely used in all the sciences—methods that look for patterns in numerical data and display those patterns graphically. The MDS in figure 16.3 shows the pattern in table 16.2. That is, it shows how one informant sees the similarities among the 18 fruits.

Look carefully at table 16.2. There are 1s in the cells 2-11, 2-13, and 11-13. This is because the informant put orange (2), lemon (11), and grapefruit (13) in one pile and nothing else in that pile. This behavior is presented graphically in figure 16.3 with the orange-lemon-grapefruit cluster shown separated from other clusters.

How MDS Works

You'll sometimes see **multidimensional scaling** called **smallest-space analysis.** That's because MDS programs work out the best spatial representation of a set of objects that are represented by a set of similarities. Suppose, for example, that you measure the distance, in miles, among three cities, A, B, and C. The matrix for these cities is in the inside box of table 16.3.

Clearly, cities A and C are closer to one another than are A and B, or B and C. You can represent this with a triangle, as in figure 16.4a.

In other words, we can place points A, B, and C on a plane in some position relative to each other. The distance between A and B is longer than that between A and C (reflecting the difference between 40 and 50 miles); and the distance between B and C is longer than that between A and C (reflecting the difference between 40 and 80 miles).

A rule in graph theory says that you can plot the relations among any set of relations

Rule 1

IF Child has blood stools OR child has swollen glands OR child is vomiting
THEN take child to doctor.

Rule 2

IF diarrhea is caused by *empacho*
THEN give physical treatment.

Rule 3

IF previous rules do not apply OR there is no cure with *empacho* treatment
THEN give the highest preferred curing treatment that meets constraints.

Rule 4

IF previous treatment did not stop diarrhea
THEN compare the two highest preferred treatments of remaining options.

Rule 4.1

IF one is a curing remedy AND meets its constraints
THEN give this treatment.

Rule 4.2

IF both or neither are curing remedies AND each meets its respective constraints
THEN give the highest-ranked preference.

Rule 5

IF the previous treatment did not stop the diarrhea AND
the episode is less than 1 week
THEN repeat rule 4.

Rule 6

IF the episode has lasted more than 1 week
THEN take the child to a doctor.

Constraints

IF you know how to make ORS (oral rehydration solution) AND
your child will drink ORS
THEN give ORS.
IF you know a medication that works for diarrhea AND
you have it in the house
THEN give the pill or liquid medication.
IF you know a medication that works for diarrhea AND
it is cheap AND
it is easy to obtain
THEN give the pill or liquid medication.

FIGURE 17.6.
Ryan and Martínez's decision model as a series of IF-THEN rules.
SOURCE: G. W. Ryan and H. Martínez, "Can We Predict What Mothers Do? Modeling Childhood Diarrhea in Rural Mexico," *Human Organization*, Vol. 55, pp. 47–57, 1996. Reprinted with permission of the Society for Applied Anthropology.

FOLK TAXONOMIES

There are about 6,000 languages spoken in the world today. Speakers of all those languages name things in the natural world. In 1914, Henderson and Harrington published a monograph on the ethnozoology of the Tewa Indians of New Mexico. Scholars ever since have been interested in understanding the variety of ways in which people organize their knowledge of the natural world.

In the 1950s, anthropologists began systematically producing **folk taxonomies**—that is, hierarchical, taxonomic graphs to represent how people organize their knowledge of plants and animals. These ethnobotanical and ethnozoological taxonomies don't necessarily mirror scientific taxonomies, but the whole point of what became known as **ethnoscience** is to understand cultural knowledge on its own terms.

Scientific taxonomies for plants and animals recognize six primary levels of distinction (phylum, class, order, family, genus, and species) and lots of in-between levels as well (infraorder, superorder, subclass, etc., etc.), but folk taxonomies of plants and animals across the world are generally limited to five or, at most, six levels. Figure 17.7 (from D'Andrade 1995) shows part of the folk taxonomy of *creatures* for native speakers of English.

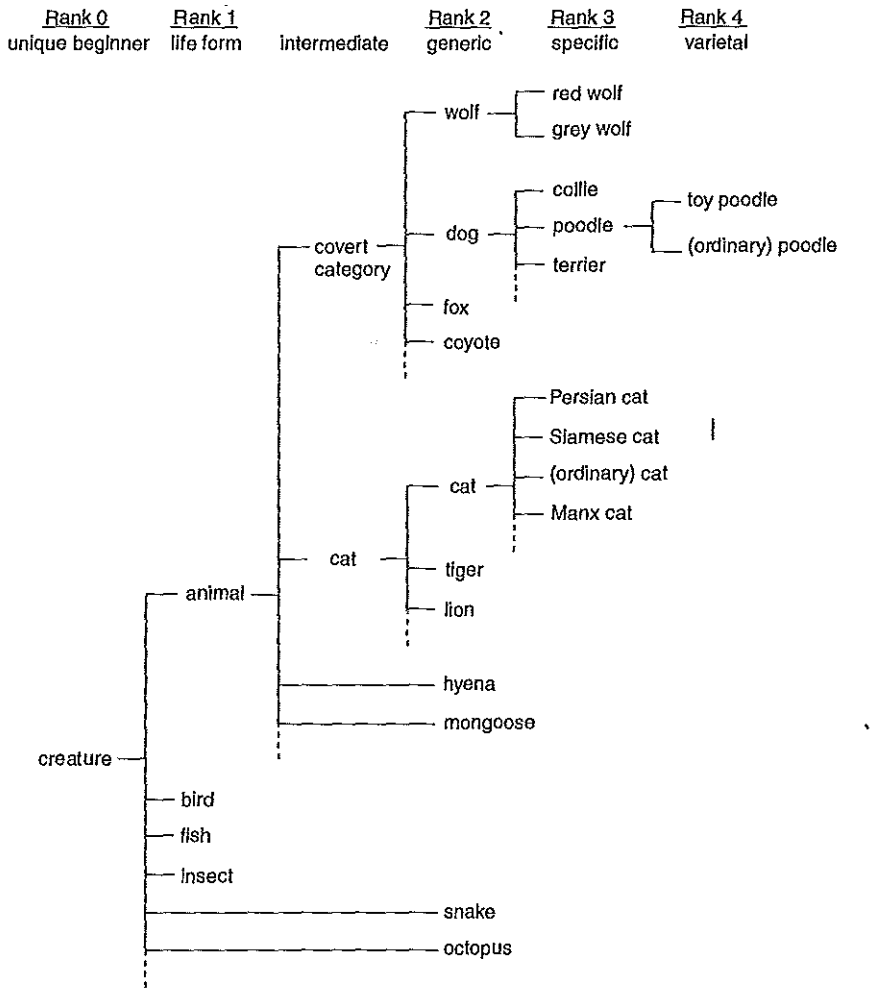

FIGURE 17.7.
Partial taxonomy for *creatures* in English.
SOURCE: R. G. D'Andrade, *The Development of Cognitive Anthropology*, p. 99. © 1995 Cambridge University Press. Reprinted by permission.

Covert Categories

There are six culturally appropriate levels of hierarchical distinction identified in figure 17.7: (1) First, there is the unique beginner, a single label that identifies the cultural domain. (2) Next comes a relatively small set of life forms (animals, fish, insects, etc.). (3) Then there is an intermediate level, which includes **covert categories**, if any exist in a particular taxonomy. Folk genera (level 4), folk species (level 5), and folk varieties (level 6) round out the picture.

There is a covert, unnamed category in figure 17.7 comprising wolves, foxes, dogs, coyotes, and some other things (the dashed line extending down from coyote, indicates that the covert category contains more than what's listed in the figure). In a scientific taxonomy, foxes are not in the same genus with dogs and wolves. The latter are in the genus *Canis*, while foxes are in the genus *Vulpes*. Many speakers of English, however, classify foxes and wolves in the category of "things in the dog family," or "canines," and a folk taxonomy of English animal terms respects that.

The intermediate category of "cat" is not covert. How can you tell? As D'Andrade says, you can say "Look at that cat!" if you're talking about a tiger, but it's weird to say "Look at that dog" if you're pointing to a fox, so "cat" is a named intermediate category and "dog" isn't.

Two more things about figure 17.7. Note how we use words for generic animals in English that would be at the species level in a scientific taxonomy (wolf, coyote, and dog are all members of the genus *Canis*, species *lupus*, *latrans*, and *familiaris*, respectively), and how the species level in the folk taxonomy comprises names for subspecies in a scientific taxonomy.

Also, look at how D'Andrade has placed octopus and snake and snake in figure 17.7. The horizontal lines show that D'Andrade has classified these creatures as **nonaffiliated generics**. They might be classified as life forms, but, as D'Andrade points out, there are many nonaffiliated generics in the ocean, including clams, lobsters, seahorses, jellyfish, and octopi.

Cultural Domains and Folk Taxonomies

It was quickly recognized that folk taxonomies could be developed for *any* **cultural domain**, not just for ethnobotanical and ethnozoological knowledge, and that we use folk taxonomies all the time to order our experience and guide our behavior.

Take someone to a supermarket—one they've never been to before—and ask them to find peanut butter. Follow them as they make their way around the store and get them to talk about what they think they're doing. Here's a typical response:

> Well, let's see, milk and eggs are over there by that wall, and the meat's usually next to that, and the canned goods are kind of in the middle, with the soaps and paper towels and stuff on the other side, so we'll go right in here, in the middle. No, this is the soap aisle, so let's go over to the right. . . . Sure, here's the coffee, so it's got to be on this aisle or the next, with cans of things like ravioli.

Any competent member of U.S. or Canadian culture will find the peanut butter in a hurry, but not everything is so clear. Shredded coconut and walnuts are often shelved with flour in the United States because they are used in baking, but other nuts—cashews and peanuts, for example—may be shelved somewhere else, like with the snacks. Lychee nuts (a Chinese dessert food) and matzohs (unleavened bread boards eaten primarily by Jews) are sometimes shelved in U.S. supermarkets together under "ethnic foods," but may

be shelved in separate "Oriental foods" and "Jewish foods" sections if local populations of those groups are sufficiently large.

How to Make a Taxonomy: Pile Sorts

Pile sorting is an efficient method for generating taxonomic trees (Werner and Fenton 1973). Simply hand informants the familiar pack of cards, each of which contains some term in a cultural domain. Informants sort the cards into piles, according to whatever criterion makes sense to them. After the first sorting, informants are handed each pile and asked to go through the exercise again. They keep doing this until they say that they cannot subdivide piles any further. At each sorting level, informants are asked if there is a word or phrase that describes each pile.

Perchonock and Werner (1969) used this technique in their study of Navajo animal categories. After an informant finished doing a pile sort of animal terms, Perchonock and Werner built a branching tree diagram, like the one in figure 17.8. They would ask the informant to make up sentences or phrases that expressed some relation between the nodes. They found that informants intuitively grasped the idea of tree representations for taxonomies.

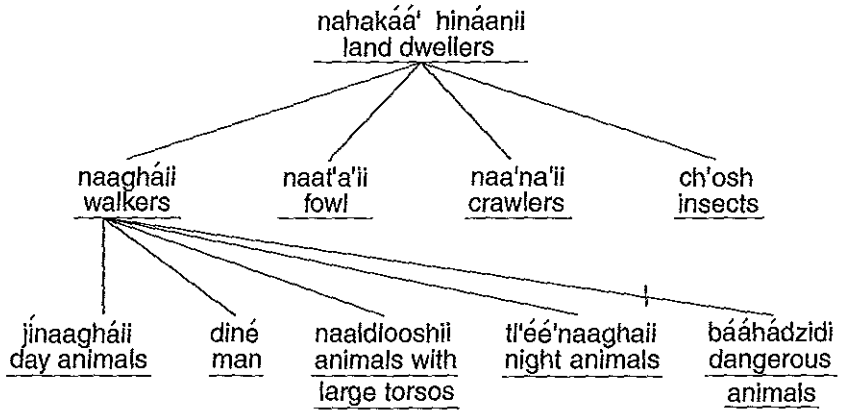

FIGURE 17.8.
Part of the Navajo animal kingdom, derived from a pile sort.
SOURCE: N. Perchonock and O. Werner, "Navaho Systems of Classification: Some Implications for Ethnoscience." *Ethnology*, Vol. 8, pp. 229–42. Copyright © 1969. Reprinted by permission.

How to Make a Taxonomy: Lists and Frames

In building a folk taxonomy, many researchers combine the free-list and frame elicitation techniques I described in chapter 10. Start with the frame:

What kinds of _____ are there?

where the blank is "cars," "trees," "saddles," "snow," "soldiers"—whatever you're interested in understanding. This frame is used again and again, until an informant says that the question is silly.

For example, suppose you asked a native speaker of American English "What kinds of foods are there?" You might get a list like: pasta, meat, fish, fruits, vegetables, snacks. . . ."

(You'll probably get a slightly different set of labels if you ask a native speaker of British English this same question.)

Next, you ask: "What kinds of pasta [meats] [fish] [etc.] are there?" The answer for meats might be: beef, lamb, chicken, pork, venison. . . .

So you extend the search: "What kinds of beef [lamb] [chicken] [etc.] are there?" For some people, at least, you'll find that beef is divided into steak, chops, hamburger, and so on, and that chicken is divided into dark meat and white meat. But if you ask "What kinds of steaks are there?" you might be told: "There are no kinds; they just are what they are." If you're dealing with a real steak lover, you might be told about Porterhouse, T-bone, rib eye, Delmonico, filet mignon, and so on.

Once you have a list of lexical items in a domain, and once you've got the basic divisions down, the next step is to find out about overlaps. Some foods, like peanuts, get classified as snacks and as protein sources by different people—or even by the same person at different times.

The point is, although the Food and Drug Administration may have codified foods in the United States, there is no codified set of folk rules for a taxonomy of foods in U.S. culture. The only way to map this is to construct folk taxonomies from information provided by a number of people and to get an idea of the range of variation and areas of consistency in how people think about this domain. You can learn about the possible overlaps in folk categories by using the substitution frames:

Is _____ a kind of _____ ?
Is _____ a part of _____ ?

Once you have a list of terms in a domain, and a list of categories, you can use this substitution frame for all possible combinations. Are marshmallows a kind of meat? A kind of fish? A kind of snack? This can get really tedious, but discovering **levels of contrast**—that *magenta* is a kind of *red*, that *cashews* are a kind of *nut*, that *alto* is a kind of *sax*, or that *ice cream* is a kind of *dessert*—just takes plain hard work. Unless you're a child, in which case all this discovery is just plain fun.

A common way to display folk taxonomies is with a branching tree diagram. Figure 17.9 shows a tree diagram for part of a folk taxonomy of passenger cars. I elicited this taxonomy in Morgantown, West Virginia, from Jack in 1976.

Things to Look for in Folk Taxonomies

There are five points to make about the taxonomy shown in figure 17.9:

1. Interinformant variation is common in folk taxonomies. That is, different people may use different words to refer to the same category of things. Sometimes, in fact, terms can be almost idiosyncratic. Jack distinguished among what he called "regular cars," "station wagons," and "vans." The term "regular cars" is not one your normally see in automobile ads, or hear from a salesperson on a car lot.
2. Category labels do not necessarily have to be simple lexical items, but may be complex phrases. The category labeled "4-wheel drive" vehicles in figure 17.9 was sometimes called "off-road vehicles" in 1976, or even "vehicles you can go camping in or tow a horse trailer with." Jack said that Jeep station wagons were both wagons *and* 4-wheel-drive cars you can go camping in.
3. Labels change over time. By the 1990s, those cars that Jack had called "vehicles you can go camping in or tow a horse trailer with" were being called "utes" by some

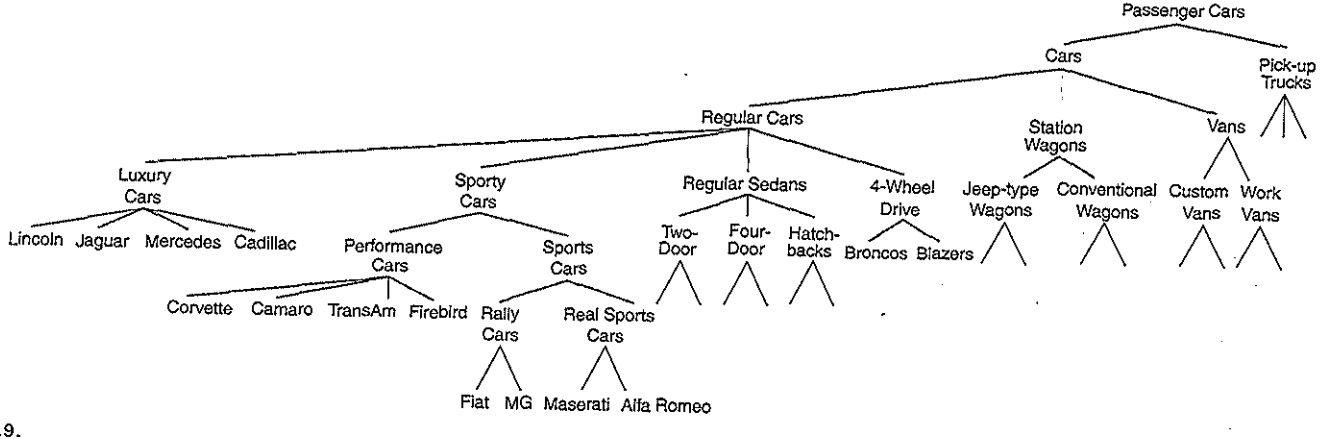

FIGURE 17.9.
Part of Jack's taxonomy of cars and trucks.

people—short for "sport utility vehicle." Today, the term widely used is SUV, though small SUVs are sometimes called "cute utes."

4. There are those covert categories I mentioned—categories for which people have no label at all, or at least not one they find easily accessible. Some people insist that Corvettes, Camaros, Maseratis, and MGs are part of a single category, which they find difficult to name (one informant suggested "sporty cars" as a label). Others, like Jack, separate "performance cars" from "sports cars" and even subdivide sports cars into "true sports cars" and "rally cars." Be on the lookout for unlabeled categories (that is, unlabeled nodes in a branching tree diagram) in any folk taxonomy.
5. Even when there are consistent labels for categories, the categories may represent multiple dimensions, each of which has its own levels of contrast. For example, many native speakers of American English recognize a category of "foreign cars" that cuts across the taxonomy in figure 17.9. There are foreign sports cars, foreign luxury cars, and foreign regular cars.

Folk taxonomies can be very, very complex. One way to get at the complexity is through multidimensional scaling (see chapter 16). Another is a technique known as componential analysis (**Further Reading:** folk taxonomies).

COMPONENTIAL ANALYSIS

Componential analysis is a formal, qualitative technique for studying meaning. There are two objectives: (1) to specify the conditions under which a native speaker of a language will call something (like a plant, a kinsman, a car) by a particular term and (2) to understand the cognitive process by which native speakers decide which of several possible terms they should apply to a particular thing.

The first objective is descriptive, but the second is a kind of **causal analysis** and is what the developers of the technique had in mind in the 1950s and 1960s (see Conklin 1955; Frake 1962; Goodenough 1956; Wallace 1962). Charles Frake, for example, described componential analysis as a step toward "the analysis of terminological systems in a way which reveals the conceptual principles that generate them" (1962:74). This created a lot of criticism, but more on that later.

Componential analysis is based on the principle of **distinctive features in phonology**, the branch of linguistics devoted to the study of the sounds of a language. To understand the principle, think about the difference in the sounds represented by P and B in English. Both are made by twisting your mouth into the same shape. This is a *feature* of the P and B sounds called "bilabial" or "two-lipped."

Another feature is that they are both "stops." That is, they are made by stopping the flow of air for an instant as it moves up from your lungs and releasing the flow suddenly. An S sound, by contrast, also requires that you restrict the air flow, but not completely. You kind of let the air slip by in a hiss. The only difference between a P and a B sound is that the P is voiceless while the B is voiced—you vibrate your vocal cords while making a P.

If you add up all the phonological features of the words "bit" and "pit," the only feature that differentiates them is voicing on the first sound in each word. The "pitness" of a pit and the "bitness" of a bit are clearly not in the voicelessness or voicedness of the sounds P and B, but any native speaker of English will distinguish the two words, and their meanings, and can trace the difference between them to that little feature of voicing if you push them a bit.

There is a unique little bundle of features that define each of the consonantal sounds