8

Nonprobability Sampling and Choosing Informants

If your objective is to estimate a parameter or a proportion from a sample to a larger population, and if your research calls for the collection of data about attributes of individuals (whether those individuals are people or organizations or episodes of a sitcom), then the rule is simple: Collect data from a sufficiently large, randomly selected, unbiased sample. If you know that you *ought* to use an unbiased sample, and you have the means to *get* an unbiased sample, and you still choose to use a nonprobability sample, then expect to take a lot of flak.

There are, however, three quite different circumstances under which a nonprobability sample is exactly what is called for:

1. Nonprobability samples are always appropriate for labor-intensive, in-depth studies of a few cases. Most studies of narratives are based on fewer than 50 cases, so every case has to count. This means choosing cases on purpose, not randomly. In-depth research on sensitive topics requires nonprobability sampling. It can take months of participant observation fieldwork before you can collect narratives about topics like sexual and reproductive history or bad experiences with mental illness or use of illegal drugs.

Come to think of it, just about everything is a sensitive topic when you dig deeply enough. Sexual history is an obviously sensitive topic, but so is the management of household finances when you get into how people really allocate their resources. People love to talk about their lives, but when you get into the details of a life history, you quickly touch a lot a nerves. Really in-

depth research requires informed informants, not just responsive respondents—that is, people whom you choose on purpose, not randomly.

2. Nonprobability samples are also appropriate for large surveys when, despite our best efforts, we just can't get a probability sample. In these cases, use a nonprobability sample and *document the bias*. That's all there is to it. No need to agonize about it.
3. And, as I said at the beginning of chapter 6, when you are collecting cultural data, as contrasted with data about individuals, then expert informants, not randomly selected respondents, are what you really need. Think of the difference between asking someone "How old was your child when you first gave him an egg to eat?" versus "At what age do children here first eat eggs?" I deal with the problem of selecting cultural experts (people who are likely to really know when most mothers introduce eggs around here) in the second part of this chapter.

The major nonprobability sampling methods are: **quota sampling**, **purposive** or (**judgment**) **sampling**, **convenience** (or **haphazard**) **sampling**, and **chain referral** (including **snowball** and **respondent-driven**) **sampling**. Finally, **case control sampling** combines elements of probability and nonprobability sampling.

Quota Sampling

In quota sampling, you decide on the subpopulations of interest and on the proportions of those subpopulations in the final sample. If you are going to take a sample of 400 adults in a small town in Japan, you might decide that, because gender is of interest to you as an independent variable, and because women make up about half the population, then half your sample should be women and half should be men. Moreover, you decide that half of each gender quota should be older than 40 and half should be younger; and that half of each of those quotas should be self-employed and half should be salaried.

When you are all through designing your quota sample, you go out and fill the quotas. You look for, say, five self-employed women who are over 40 years of age and who earn more than 300,000 yen a month and for five salaried men who are under 40 and who earn less than 300,000 yen a month. And so on.

Tinsley et al. (2002) interviewed 437 elderly users of Lincoln Park in Chicago. They selected quota samples of about 50 men and 50 women from each of the four major ethnic groups in the area, Blacks, Whites, Hispanics, and Asian Americans. Besides gender and ethnicity, Tinsley et al. stratified on place and time. They divided the park into three zones (north, south, and middle) and three time periods (6 A.M. to 10 A.M., 11 A.M. to 3 P.M., and 4 P.M. to 8 P.M.). There were, then, nine zone-time strata in which interviewers selected

respondents. The interviewers were also told to make sure they got some weekday and some weekend users of the park.

Commercial polling companies use quota samples that are fine-tuned on the basis of decades of research (Weinberger 1973). Organizations like Gallup, Roper, Harris, and others have learned how to train interviewers not to choose respondents who are pretty much like themselves; not to select only people whom they would enjoy interviewing; not to avoid people whom they would find obnoxious or hostile; not to avoid people who are hard to contact (busy people who are hardly ever home, or people who work nights and sleep days); and not to favor people who are eager to be interviewed.

The result is quota samples that are *not unbiased* but that often do a good job of reflecting the population parameters of interest. In other words, quota sampling is an art that often approximates the results of probability sampling at less cost and less hassle than strict probability sampling.

Often, but not always. In 1948, pollsters predicted, on the basis of quota sampling, that Thomas Dewey would beat Harry Truman in the U.S. presidential election. The *Chicago Tribune* was so confident in those predictions that they printed an edition announcing Dewey's victory—while the votes were being counted that would make Truman president.

Skip to 1992. In the general election in Britain that year, four different polls published on the day of the election put the Liberal Party, on average, about 1 point ahead of the Conservative Party. All the polls were based on quota sampling. The Conservatives won by 8 points. In fact, from 1992 to 1997, political polls using quota samples in Britain systematically overestimated the support for the Liberals (Curtice and Sparrow 1997).

Quota samples are biased toward people you can find easily. This means that quota sampling is simply dangerous when it comes to making predictions about election outcomes—or estimating any population parameter, for that matter.

On the other hand, quota sampling is appropriate in the study of cultural domains. If you want to know how junior sports—Little League Baseball, Pop Warner football, Youth Soccer, junior and senior high school football—function in small communities across the United States, you'd ask people who have children playing those sports. There will be some intracultural variation, but open-ended interviews with four or five really knowledgeable people will produce the relevant cultural data—including data on the range of ideas that people have about these institutions.

Many studies of narratives are based on small samples, simply because there is so much work involved. If you are doing narrative analysis, set up a quota sampling design. First, figure out how many narratives you can collect, transcribe, and code for themes. More about all this in chapter 17 on text anal-

ysis, but narratives, like life histories, can take several interviews and many hours just to collect. Figure on 6–8 hours to transcribe each recorded hour when you start out; you'll cut the time in half as you get better at it, assuming you have decent typing skills, and you'll cut it in half again if you're willing to spend the time it takes to train voice recognition software—see chapter 9. And when you get through transcribing, there's still coding to do (more hours) and analysis and write up (lots more hours).

Suppose you think you can do 40 in-depth interviews. That means you can have up to three independent, binary variables if you want five cases in each cell, and that tells you what your sampling design will look like. Suppose you are studying, through narratives, the lived experiences of labor migrants to the United States who are back home in their community in Mexico. You decide that you want to compare the experiences of (1) people who spent time on jobs in the United States but were caught by the U.S. Border Patrol and deported with those of (2) people who managed to stay on their jobs until they'd accumulated enough money to return on their own. You also want to compare the experiences of Indians and mestizos and of men and women. That's three binary independent variables: deported/not deported; Indian/mestizo; male/female. There are, as you can see from table 8.1, 16 cells in this

TABLE 8.1
Sampling Design for Three Dichotomous Variables

Mexicans who have worked in the United States and returned home							
Caught by the Border Patrol and deported							
Indian		Mestizo		Indian		Mestizo	
Male	Female	Male	Female	Male	Female	Male	Female
Returned home on their own							
Indian		Mestizo		Indian		Mestizo	
Male	Female	Male	Female	Male	Female	Male	Female

design. If you want to compare more variables simultaneously, you'll have more cells on the bottom row and you'll need more data.

Purposive or Judgment Sampling

In purposive sampling, you decide the purpose you want informants (or communities) to serve, and you go out to find some. This is somewhat like

quota sampling, except that there is no overall sampling design that tells you how many of each type of informant you need for a study. You take what you can get.

I used purposive sampling in my study of the Kalymnian (Greek) sponge-fishing industry (1987). I knew I had to interview sponge merchants, boat owners, and divers, but my first interviews taught me that I had to interview people whom I had never considered: men who used to be divers but who had quit, gone to Australia as labor migrants, and returned to their island. It was very easy to find those returned migrants: Everyone on the island either had one in their family or knew people who did.

There are many good reasons for using purposive samples. They are used widely in (1) **pilot studies**, (2) **intensive case** studies, (3) **critical case studies**, and (4) studies of **hard-to-find populations**.

1. Pilot studies. These are studies done before running a larger study. In 1999, Katherine Browne, Carla Freeman, and Zobeida Bonilla began a comparative ethnographic study of women entrepreneurs in Martinique, Barbados, and Puerto Rico—that is, in the French-, English-, and Spanish-speaking Caribbean. In a large, multisite study like this, it pays to spend time on pilot research. Each member of the team did 30 in-depth interviews with women who were engaged in a wide range of enterprises, who were of different ages, and who came from one- and two-parent homes. This helped the team develop their research instruments and provided the baseline for the larger project (Browne 2001).

And speaking of instruments, when you do surveys to test hypotheses, you want to make sure that you test all your scales with a pilot sample. More about all this in chapter 12.

2. In intensive case studies, the object is often to identify and describe a cultural phenomenon. Dickerson et al. (2000) studied the experiences of American Indian graduate nursing students and cultural barriers that might lead the students to drop out of their training. Dickerson et al. found and interviewed 11 students who were enrolled in an advanced nurse practitioner program. Samples don't get much more purposive than this, and they don't get much more appropriate, either.

Life history research and qualitative research on special populations (drug addicts, trial lawyers, shamans) rely on judgment sampling. Barroso (1997), for example, studied a purposive sample of 14 men and 6 women in the Tampa, Florida, area, all of whom had lived with AIDS for at least 3 years.

Finally, researchers don't usually pull research sites—villages, tribal encampments, hospitals, school systems—out of a hat. They rely on their judgment to find one that reflects the things they are interested in.

3. Critical case studies. Polling companies try to identify communities across the United States that have voted for the winner in the past, say, six presidential elections. Then they poll those few communities that meet the criterion.

Choosing key informants in ethnographic research is also critical-case sampling. It would be pointless to select a handful of people randomly from a population and try to turn them into trusted key informants.

4. We almost always have to rely on purposive sampling in the study of hard-to-find populations.

Think about locating and interviewing refugees from Somalia and Ethiopia living in a large American city. Many of these people experienced torture and don't exactly welcome researchers who want to ask them a lot of questions. This was the problem facing researchers in Minneapolis (see Spring et al. 2003; Jaranson et al. 2004). The study design called for a quota sample of 1,200 respondents, including 300 Oromo women, 300 Oromo men, 300 Somali women, and 300 Somali men. The study team recruited male and female interviewers from the community—people who shared ethnicity, language, and religion with the people they were trying to locate and interview. The project team sent out fliers, placed announcements in church bulletins, and made presentations at meetings of Oromo and Somali organizations. The interviewers also used their own social networks to locate potential respondents. Over 25 months, the team built trust in the community and wound up with 1,134 of the 1,200 interviews called for in the study.

Kimberly Mahaffy (1996) was interested in how lesbian Christians deal with the cognitive dissonance that comes from being rejected by mainstream Christian churches. Mahaffy sent letters to gay Christian organizations, asking them to put an ad for potential respondents in their newsletters. She sent flyers to women's bookstores and to lesbian support groups, asking for potential respondents to get in touch with her.

Eventually, Mahaffy got 163 completed questionnaires from women who fit the criteria she had established for her research, including 44 from women who self-identified as born-again or evangelical Christians. Mahaffy could not possibly have gotten an unbiased sample of lesbian Christians, but the corpus of data that she collected from her respondents had all the information she needed to answer her research questions.

Convenience or Haphazard Sampling

Convenience sampling is a glorified term for grabbing whoever will stand still long enough to answer your questions. Sometimes, convenience samples

are all that's available, and you just have to make do. Studies of the homeless, for example, are usually done with convenience samples, for obvious reasons, as are studies of people who are in intensive care units in hospitals. All samples represent *something*. The trick is to make them representative of what *you* want them to be. That's what turns a convenience sample into a purposive one.

For example, Al-Krenawi and Wiesel-Lev (1999) wanted to understand the emotions of Israeli Bedouin women who had experienced genital mutilation. They interviewed a convenience sample of 12 women who had been through the ritual and 12 women who had not but had either seen it first-hand or had heard about women in their own extended families going through it. We wouldn't put much stock in the fact that a specific *percentage* of the women reported sexual problems or relationship problems with various members of their family, but the *list* of problems is very instructive because it is the basis for more in-depth research.

If you want to estimate a parameter, then you know what you have to do: get a random, representative sample. If you want to know the percentage of adult men in a matrilateral, cross-cousin society who have actually married their biological mother's-brother's-sister (MBZ), you'll either have to count them all or take a random, unbiased sample of sufficient size to be able to make that generalization.

Key informants will tell you that the rule is broken regularly, but not by how much. A convenience sample of women who gather at the village well each day will tell you the range of options for men who don't have a biological MBZ, but not how many choose each option. And if you want to know the effect of a new road on some peasants and you only interview people who come to town on the road, you'll miss all the people who live too far off the road for it to do them any good.

Chain Referral, or Network Sampling: The Snowball and RDS Methods

Snowball and **respondent-driven sampling** (**RDS**) are two network sampling methods (also known, generically, as **chain referral** methods) for studying hard-to-find or hard-to-study populations. Populations can be hard to find and study for three reasons: (1) they contain very few members who are scattered over a large area (think strict vegans in rural Georgia); and/or (2) they are stigmatized and reclusive (HIV-positive people who never show up at clinics until they are sick with AIDS) or even actively hiding (intravenous drug users, for example); and/or (3) they are members of an elite group and don't care about your need for data.

In the snowball technique, you use key informants and/or documents to locate one or two people in a population. Then, you ask those people to (1) list others in the population and (2) recommend someone from the list whom you might interview. You get handed from informant to informant and the sampling frame grows with each interview. Eventually, the sampling frame becomes saturated—that is, no new names are offered.

David Griffith and his colleagues used *two* snowball samples in their study of food preferences in Moberly, Missouri. They chose an initial "seed" household in a middle-income neighborhood and asked a man in the house to name three people in town with whom he interacted on a regular basis. The first person cited by the informant lived in a lower-income neighborhood across town. That person, in turn, named other people who were in the lower-income bracket.

After a while, the researchers realized that, though they'd started with a middle-income informant who had children at home, they were getting mostly lower-income, elderly people in the snowball sample. So they started again, this time with a seed from an elite, upper-middle-income neighborhood. By the time they got through, Griffith et al. had a well-balanced sample of 30 informants with whom they did in-depth interviews (reported in Johnson 1990:78).

Thomas Weisner has been following 205 counterculture women and their families since 1974. Weisner built this sample by recruiting women in California who were in their third trimester of pregnancy. He used snowball sampling, but to ensure that participants came from all over the state and represented various kinds of families, he used no more than two referrals from any one source (Weisner 2002:277).

Snowball sampling is popular and fun to do, but in large populations it does not produce a random, representative sample. If you are dealing with a relatively small population of people who are likely to be in contact with one another, like practitioners of alternative medicine in a small town, then snowball sampling is an effective way to build an exhaustive sampling frame. Once you have an exhaustive sampling frame, you can select people at random to interview. In this case, snowball sampling is one step in a two-step process for getting a representative sample.

For large populations, however, people who are well known have a better chance of being named in a snowball procedure than are people who are less well known. And in large populations, people who have large networks name more people than do people who have small networks. For large populations, then, snowball sampling is risky because every person does not have the same chance of being included.

Douglas Heckathorn (1997) developed respondent-driven sampling for

dealing with these problems. Like snowball sampling, RDS begins with a few informants who act as seeds. The informants are paid for being interviewed and are then asked to recruit up to three members of their networks into the study. To move this process along, Heckathorn paid each of his see informants $10 and gave them three coupons. Anyone who came to Heckathorn to be interviewed and who had one of those coupons was paid the same $10. (He upped the bounty to $15 for referring a female drug injector, since they were harder to find.) Those informants, in turn, got several coupons and recruited others into the study.

There are several very important improvements to snowball sampling here. First, this method avoids the ethical problem that snowball sampling presents. The people whom an informant names may not want you even to know about their existence, much less be anxious to grant you an interview. Second, having members of a hard-to-study population do the recruiting deals with the reluctance of some people to be interviewed. And finally, Heckathorn (1997, 2002) shows that, when it's done right, the RDS method produces samples that are less biased than are traditional snowball samples. (For more on chain referral sampling, see Sudman and Kalton 1986, Martin and Dean 1993, Heckathorn and Jeffri 2001, and Salganik and Heckathorn 2004.)

Case Control Sampling

In case control sampling, you choose a purposive sample on the basis of some criterion (like having a certain illness or injury, or attempting suicide, or being homeless) and match the members of that sample with people who match the cases on many criteria, but *not* on the case criterion. This method is widely used in public health research.

For example, Beautrais et al. (1998) had a set of 129 cases of men and women under 25 who had survived a medically serious attempt at suicide in Christchurch, New Zealand. (A medically serious attempt at suicide generally involves treatment in a hospital for more than 24 hours.) Beautrais et al. recruited 153 control cases from the electoral rolls of the area, such that the mean age of the controls matched the mean age of the attempted suicide cases and the proportion of men and women in the two samples were more or less the same. In this study, both the study cases and the control cases were purposive.

Case control sampling has great potential for field research as well. Bassuk and Rosenberg (1988) wanted to know why whole families in Boston were homeless. By canvasssing homeless shelters, they identified 49 female-headed, homeless families with a total of 86 children. Then they selected 81

control cases—poor, female-headed families in the same city. They made sure that the two samples—the cases and the controls—were about the same in average age of the mother (28 and 29), average age when the mother had her first child (20 and 19), and average number of children (2.4 and 2.5). Then these researchers looked for *differences* between the cases and the controls to see what might account for homelessness.

And case control is not limited to places like Christchurch and Boston. Pfeiffer et al. (2001) used the method to study why some Shona children, in Sussundenga, Mozambique, are malnourished, while others thrive. Mothers of children under 5 years of age in Sussundenga brought their children in regularly to a local clinic. Using a table like that in appendix A, Pfeiffer et al. took a random sample of 50 undernourished children (no more than one from a household) from the clinic's records. These are the **index cases**.

Next, they walked away from each of the index homes in a random direction (you just select a random number between 1 and 4) and visited every house along that path until they found the first suitable **control case**. They looked for households with: (1) one or more children under five, none of whom were malnourished; (2) more or less the same income as one of the index-case households; and (3) the same house construction (same kind of latrine, same roof material), and the same water access as one of the index-case households. The control case was the oldest child under 5 in the control household.

When they started the research, Pfeiffer et al. hypothesized that thriving children would be in households where women brought in more money (or managed more money) than women do in index-case households. The hypothesis was not supported. Instead, control mothers had about double the average education of case mothers and mothers in the control households reported more than twice the number of protein consumption days in the past month (meat, fish, poultry) as did mothers in the index households.

Sampling and Credibility

Particularly in ethnographic research, you learn in the field, as you go along, to select the units of analysis (people, court records, whatever) that will provide the information you need. This is what Russell Belk et al. (1988) did in their detailed ethnographic study of buyers and sellers at a swap meet. When you study a process, like bargaining over goods, and you're doing the research in the field, in real time (not under simulated conditions in a lab), then selecting informants who meet certain criteria is the right thing to do.

The credibility of research results comes from the power of the methods used in measurement and sampling. Good measurement is the key to internal

validity and representative sampling is the key to external validity. Well-done nonprobability sampling is actually part of good measurement. It contributes to credibility by contributing to internal validity. When someone reads a research report based on really good measurement of a nonprobability sample, they come away thinking, "Yep, I believe those conclusions about the people who were studied in that piece of research."

That's plenty. If you want the credibility of your conclusions to extend beyond the group of people (or countries, or organizations, or comic books) you studied, then either: (1) Repeat the study one or more times with nonprobability samples; or (2) Use a probability sample.

Choosing Informants

Across the social sciences, you'll see references to research participants as "respondents," or "subjects," or "informants." These terms tend to be used by sociologists, psychologists, and anthropologists, respectively. Respondents respond to survey questions, subjects are the subject of some experiment, and informants . . . well, informants tell you *what they think you need to know* about their culture.

There are two kinds of informants: **key informants** and **specialized informants**. Key informants are people who know a lot about their culture and are, for reasons of their own, willing to share all their knowledge with you. When you do long-term ethnography you develop close relationships with a few key informants—relationships that can last a lifetime. You don't choose these people. They and you choose each other, over time.

Specialized informants have particular competence in some cultural domain. If you want to know the rules of Balinese cockfighting, or how many cows must be given to a Lumasaba bride's parents, or when to genuflect in a Roman Catholic Mass, or what herb tea to give children for diarrhea, you need to talk to people who can speak knowledgeably about those things.

Key Informants

Good key informants are people whom you can talk to easily, who understand the information you need, and who are glad to give it to you or get it for you. Pelto and Pelto (1978:72) advocate training informants "to conceptualize cultural data in the frame of reference" that you, the researcher, use.

In some cases, you may want to just listen. But when you run into a really great informant, I see no reason to hold back. Teach the informant about the

analytic categories you're developing and ask whether the categories are correct. In other words, encourage the informant to become the ethnographer.

I've worked with Jesús Salinas since 1962. In 1971, I was about to write an ethnography of his culture, the Ñähñu of central Mexico, when he mentioned that he'd be interested in writing an ethnography himself. I dropped my project and taught him to read and write Ñähñu. Over the next 15 years, Salinas produced four volumes about the Ñähñu people—volumes that I translated and from which I learned many things that I'd never have learned had I written the ethnography myself. For example, Ñähñu men engage in rhyming duels, much like the "dozens" of African Americans. I wouldn't have thought to ask about those duels because I had never witnessed one (see Bernard and Salinas Pedraza 1989).

Just as Salinas has influenced my thinking about Mexican Indian life, Salinas's ethnography was heavily influenced by his association with me. We've discussed and analyzed parts of Ñähñu culture over the years and we've even argued over interpretation of observed facts. (More about all this in the section on native ethnography in chapter 17 on text analysis, plus a different perspective by Harry Wolcott [1999].)

Finding Key Informants

One of the most famous key informants in the ethnographic literature is Doc in William Foote Whyte's *Street Corner Society* (1981 [1943]). Whyte studied "Cornerville," an Italian American neighborhood in a place he called "Eastern City." (Cornerville was the North End of Boston.) Whyte asked some social workers if they knew anyone who could help Whyte with his study. One social worker told Whyte to come to her office and meet a man whom she thought could do the job. When Whyte showed up, the social worker introduced him to Doc and then left the room. Whyte nervously explained his predicament, and Doc asked him "Do you want to see the high life or the low life?" (Whyte 1989:72).

Whyte couldn't believe his luck. He told Doc he wanted to see all he could, learn as much as possible about life in the neighborhood. Doc told him:

> Any nights you want to see anything, I'll take you around. I can take you to the joints—the gambling joints. I can take you around to the street corners. Just remember that you're my friend. That's all they need to know. I know these places and if I tell them you're my friend, nobody will bother you. You just tell me what you want to see, and we'll arrange it. . . . When you want some information, I'll ask for it, and you listen. When you want to find out their philosophy of life, I'll start an argument and get it for you. (ibid.)

Doc was straight up; he told Whyte to rely on him and to ask him anything, and Doc was good to his word all through Whyte's 3 years of fieldwork. Doc introduced Whyte to the boys on the corner; Doc hung out with Whyte and spoke up for Whyte when people questioned Whyte's presence. Doc was just spectacular.

Or was he? Boelen (1992) visited Cornerville 25 times between 1970 and 1989, sometimes for a few days, other times for several months. She tracked down and interviewed everyone she could find from *Street Corner Society*. Doc had died in 1967, but she interviewed his two sons in 1970 (then in their late teens and early 20s). She asked them what Doc's opinion of Whyte's book had been and reports the elder son saying: "My father considered the book untrue from the very beginning to the end, a total fantasy" (Boelen 1992:29).

Of course, Whyte (1996a, 1996b) refuted Boelen's report, but we'll never know the whole truth. Whyte certainly made mistakes, but the same can be said for all ethnographers. For some scholars, mistakes invalidate a positivist stance in ethnography. For others, it does not.

Doc may be famous, but he's not unique. He's not even rare. All successful ethnographers will tell you that they eventually came to rely on one or two key people in their fieldwork. What was rare about Doc is how quickly and easily Whyte teamed up with him. It's not easy to find informants like Doc. When Jeffrey Johnson began fieldwork in a North Carolina fishing community, he went to the local marine extension agent and asked for the agent's help. The agent, happy to oblige, told Johnson about a fisherman whom he thought could help Johnson get off on the right foot.

It turned out that the fisherman was a transplanted northerner; he had a pension from the Navy; he was an activist Republican in a thoroughly Democratic community; and he kept his fishing boat in an isolated moorage, far from the village harbor. He was, in fact, maximally different from the typical local fisherman. The agent had meant well, of course (Johnson 1990:56).

In fact, the first informants with whom you develop a working relationship in the field may be "deviant" members of their culture. Agar (1980b:86) reports that during his fieldwork in India, he was taken on by the *naik*, or headman of the village. The *naik*, it turned out, had *inherited* the role, but he was not respected in the village and did not preside over village meetings. This did not mean that the *naik* knew nothing about village affairs and customs; he was what Agar called a "solid insider," and yet somewhat of an outcast—a "marginal native," just like the ethnographer was trying to be (Freilich 1977). If you think about it, Agar said, you should wonder about the kind of person who would befriend an ethnographer.

In my own fieldwork (at sea, in Mexican villages, on Greek islands, in rural communities in the United States, and in modern American bureaucracies), I

have consistently found the best informants to be people who are cynical about their own culture. They may not be outcasts (in fact, they are always solid insiders), but they say they *feel* somewhat marginal to their culture, by virtue of their intellectualizing of and disenchantment with their culture. They are always observant, reflective, and articulate. In other words, they invariably have all the qualities that I would like to have myself.

Don't choose key ethnographic informants too quickly. Allow yourself to go awash in data for a while and play the field. When you have several prospects, check on their roles and statuses in the community. Be sure that the key informants you select don't prevent you from gaining access to other important informants (i.e., people who won't talk to you when they find out you're so-and-so's friend). Since good ethnography is, at its best, a good story, find trustworthy informants who are observant, reflective, and articulate—who know how to tell good stories—and stay with them. In the end, ethnographic fieldwork stands or falls on building mutually supportive relations with a few key people.

Informants Sometimes Lie

Don't be surprised if informants lie to you. Jeffrey Johnson, a skilled boat builder, worked in an Alaskan boatyard as part of his field study of a fishing community. At one point in his fieldwork, two other ethnographers showed up, both women, to conduct some interviews with the men in the boatyard. "The two anthropologists had no idea I was one of *them*," Johnson reports, "since I was dressed in carpenter's overalls, with all the official paraphernalia—hammer, tape measure, etc. I was sufficiently close to overhear the interview and, knowing the men being interviewed, recognized quite a few blatant lies. In fact, during the course of one interview, a captain would occasionally wink at me as he told a whopper of a lie" (personal communication).

This is not an isolated incident. A Comox Indian woman spent 2 hours narrating a text for Franz Boas. The text turned out to be nothing but a string of questions and answers. Boas didn't speak Comox well enough to know that he was being duped, but when he found out he noted it in his diary (Rohner 1969:61).

In 1938, Melville Herskovits published his massive, two-volume work on the ancient West African kingdom of Dahomey (today Benin). According to Herskovits, there was an annual census and the data from these efforts were used in administering the state. The counting involved the delivery of sacks of pebbles from around the kingdom to the palace at Abomey, with each pebble representing a person. Roger Sandall (1999) has shown that the informant who

told Herskovits about this elaborate accounting system may have made it all up.

This sort of thing can happen to anyone who does participant observation ethnography, but some cultures are more tolerant of lying than are others. Nachman (1984) found that the most articulate informants among the Nissan of New Guinea were great truth tellers and accomplished liars at the same time. Among the Nissan, says Nachman, people expect big men to give speeches and to "manipulate others and to create socially acceptable meanings," even if that means telling outright lies (ibid.:552).

Selecting Culturally Specialized Informants

The search for formal and systematic ways to select focused ethnographic informants—people who can help you learn about particular areas of a culture—has been going on for a very long time. In 1957, Marc-Adelard Tremblay was involved in a Cornell University survey research project on poverty in Nova Scotia. He wanted to use ethnographic informants to help the team's researchers design a useful questionnaire, so he made a list of some roles in the community he was studying—things like sawmill owners, doctors, farmers, bankers—and chose informants who could talk to him knowledgeably about things in their area of expertise. Tremblay had no external test to tell him whether the informants he selected were, in fact, the most competent in their areas of expertise, but he felt that on-the-spot clues made the selection of informants valid (Tremblay 1957).

Michael Robbins and his colleagues studied acculturation and modernization among the Baganda of Uganda, using a more formal method to select informants who might be competent on this topic (Robbins et al. 1969). First, they ran a survey of households in a rural sector, asking about things that would indicate respondents' exposure to Western culture. Then they used the results of the survey to select appropriate informants.

Robbins et al. had 80 variables in the survey that had something to do with acculturation and they ran a factor analysis to find out which variables package together. We'll look a bit more at factor analysis in chapter 21. For now, think of factor analysis as a way to reduce those 80 variables to just a handful of underlying variables around which individual variables cluster. It turned out that 14 of the original 80 variables clustered together in one factor. Among those original variables were: being under 40 years of age, drinking European beer, speaking and reading English, having a Western job, and living in a house that has concrete floors and walls.

Robbins et al. called this cluster the "acculturation factor." They chose

informants who had high scores on this factor and interviewed them about acculturation. Robbins et al. reversed Tremblay's method. Tremblay used key informants to help him build a survey instrument; Robbins et al. used a survey to find key informants.

In any given domain of culture, some people are more competent than others. In our culture, some people know a lot about the history of baseball; some people can name the actors in every sitcom since the beginning of television in the 1940s. Some people are experts on medicinal plants; others are experts on cars and trucks. John Poggie (1972) did an early study of informant competence. He selected one informant in each of seven Mexican communities. The communities ranged in size from 350 to 3,000 inhabitants. The informants were village or town presidents, or judges, or (in the case of agricultural communities) the local commissioners of communal land. Poggie asked these informants questions about life in the communities, and he compared the answers with data from a high-quality social survey.

For example, Poggie asked the seven informants: "How many men in this town are workers in Ciudad Industrial?" (Ciudad Industrial is a fictitious name of a city that attracted many labor migrants from the communities that Poggie studied.) In his survey, Poggie asked respondents if they had ever worked in Ciudad Industrial. The correlation between the answers given by Poggie's expert informants and the data obtained from the survey was .90.

Poggie also asked: "What percentage of the houses here are made of adobe?" This time, the correlation between the informants and the survey was only .71. Table 8.2 shows the seven questions Poggie asked, and how well his informants did when their answers were compared to the survey.

TABLE 8.2
Agreement between Informants and Survey Data in Seven Villages

Questions asked of informants	*Correlation with questionnaire data*
Number of men from this town who are workers in Ciudad Industrial	0.90
Percentage of houses made of adobe	0.71
Percentage of households that have radios	0.52
Percentage of people who eat eggs regularly	0.33
Percentage of people who would like to live in Ciudad Industrial	0.23
Percentage of people who eat bread daily	0.14
Percentage of people who sleep in beds	0.05

SOURCE: J. J. Poggie, "Toward Quality Control in Key Informant Data," *Human Organization*, Vol. 31, pp. 26–29, 1972. Reprinted with permission of the Society for Applied Anthropology.

Overall, informants produced answers most like those in the survey when they were asked to respond to questions about things that are publicly observable. The survey data are not necessarily more *accurate* than the informants' data. But as the questions require informants to talk about things inside people's homes (such as what percentage of people eat eggs), or about what people think (what percentage of people would *like* to live in Ciudad Industrial), informants' answers look less and less like those of the survey.

Poggie concluded that "There is little reason to believe that trust and rapport would improve the reliability and precision concerning what percentage sleep in beds, who would like to live in the new industrial city, or what percentage eat bread daily" (ibid.:29).

The Cultural Consensus Model

The idea that people can be more or less competent in various areas of their culture has led to formal tests of new methods for selecting focused ethnographic informants. James Boster (1985, 1986) walked 58 Aguaruna Jívaro women (in Peru) through a garden that had 61 varieties of manioc. He asked the women *waji mama aita*? ("What kind of manioc is this?"), and calculated the likelihood that all possible pairs of women agreed on the name of a plant. Since Boster had planted the garden himself, he *knew* the true identification of each plant. Sure enough, the more that women agreed on the identification of a plant, the more likely they were to know what the plant actually was. In other words, as cultural consensus increased, so did cultural competence.

This makes a lot of sense. Suppose you give a test about the rules of baseball to two groups of people: a group of rabid baseball fans and another group (Americans, Canadians, Mexicans, Dominicans, etc.) who never watch the game. You'd expect that: (1) The serious baseball fans will agree more among themselves about the answers to your test questions than will the nonfans; and (2) The serious fans will get the answers right more often than the nonfans. These outcomes are expected because of the relation between cultural consensus and cultural competence.

Boster's experiment and the hypothetical baseball experiment are pretty much like any test you might take in a class. The instructor makes up both the test and an answer key with the (supposedly) correct answers. Your job is to match your answers with those on the answer key.

But what if there were no answer key? That's what happens when we ask people to tell us the uses of various plants, or to list the sacred sites in a village, or to rate the social status of others in a community. We are not asking people for their opinions, attitudes, beliefs, or values. We ask informants to

rate the social status of others in their community because we want to *know* the social status of all those people. The problem is, we don't have an answer key to tell whether informants are accurate in their reporting of information.

Romney et al. (1986) developed a formal method, called the **cultural consensus model**, to test informant competence *without having an answer key*. The theory behind the technique makes three assumptions:

1. Informants share a common culture and there is a culturally correct answer to any question you ask them. The culturally correct answer might be incorrect from an outsider's perspective (as often happens when we compare folk knowledge about illnesses to biomedical knowledge). Any variation you find among informants is the result of *individual* differences in their knowledge, not the result of being members of subcultures.
2. Informants give their answers to your test questions independently of one another.
3. All the questions in your test come from the same cultural domain—that is, things that can be listed, like kinds of animals or hand tools, or things you can do on a weekend. (We'll take up cultural domain analysis in chapter 11.) A test that asks about kinship and Australian-rules football would be a poor test. People can be competent in one domain and incompetent in another. The cultural consensus method must be used *only* for identifying people who are knowledgeable about a particular domain.

To use the consensus technique, simply give a sample of informants a test that asks them to make some judgments about a list of items in a cultural domain. You can use true-false and yes-no questions. An example of a true-false question in fieldwork might be: "You can get [pneumonia] [diarrhea] [*susto*] from [being overweight] [tired] [scared] [in the room with a sick person]." Some other typical true-false questions might be: "You can't get AIDS from touching the body of someone who died from it," or "A field goal is worth 7 points."

You can also use multiple-choice questions or even open-ended, fill-in-the-blank questions. (See appendix F for information about *Anthropac*, a set of programs that includes modules for handling cultural consensus data.)

For the test to reliably distinguish cultural competence among informants, it's best to have about 40 test items and about 40 informants. As an example, table 8.3 shows the answers of four informants to a 40-question true-false test about "general knowledge" for Americans (things like who starred in some classic movies). The 1s are items to which a student answered "true" (or "yes"), and the 0s are items to which a student answered "false" (or "no").

Table 8.4 shows the *number* of matches between informants, the *proportion of matches* (the number of matches divided by the number of items in the

TABLE 8.3

Answers by Four Students to a 40-Question T/F General Knowledge Test

1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0	1	0	1	1	0	1	0	1
0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1
0	1	0	0	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	1	0	1	0	1	0	1	0	0	1	0	0
0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	0	0	1	0	0

SOURCE: "Culture as Consensus: A Theory of Culture and Informant Accuracy" by A. K. Romney et al., 1986, *American Anthropologist* 88:324. Reproduced by permission of the American Anthropological Association. Not for further reproduction.

Note: 1 represents "true" and 0 represents "false."

TABLE 8.4
Matches, Proportions of Matches, Proportions of Corrected Matches, and Competency Scores for the Data in Table 8.3

	Number of matches (matrix I)				*Proportion of matches (matrix II)*				*Proportion of corrected matches (matrix III)*				*Competency score for student*	
Informant	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*		
1	—	27	24	22	—	0.675	0.625	0.550	—	0.35	0.25	0.10	1	0.48
2	27	—	34	21	0.675	—	0.850	0.525	0.35	—	0.70	0.05	2	0.61
3	25	34	—	23	0.625	0.850	—	0.575	0.25	0.70	—	0.15	3	0.61
4	22	21	23	—	0.550	0.525	0.575	—	0.10	0.05	0.15	—	4	0.32

SOURCE: "Culture as Consensus: A Theory of Culture and Informant Accuracy" by A. K. Romney et al., 1986, *American Anthropologist* 88:324. Reproduced by permission of the American Anthropological Association. Not for further reproduction.

test), and the proportion of matches *corrected for guessing*. This correction is necessary because anyone can guess the answers to any true-false test item half the time. *Anthropac* has a built-in error-correction routine for consensus analysis.

The three matrices in table 8.4 are called **similarity matrices** because the entries in each matrix give some direct estimate of how similar any pair of informants is (see chapters 11, 16, and 21 for more on similarity matrices). Look at Matrix I, the one called "number of matches." Informants 1 and 2 have 27 matches. If you look along the first two rows of table 8.3 and count, you'll see that on 27 out of 40 test questions, informants 1 and 2 answered the same. When informant 1 said "false" (0), then informant 2 said "false" (0), and when informant 1 said "true" (1), then informant 2 said "true" (1).

Now look at Matrix II, "proportion of matches." This shows that informants 1 and 2 were 67.5% similar, because 27/40 = .675. Finally, look at Matrix III, "proportion of corrected matches." After correcting for the possibility that some of the similarity in Matrix II between informants is due to the fact that they guessed the same answers when they didn't really know the answers, we see that informants 1 and 2 are .35 alike, while informants 2 and 3 are .70 alike. Informants 2 and 3 are twice as similar to one another as informants 1 and 2 are to one another.

Look down the last column of Matrix III. Informant 4 is not like any other informant. That is, informant 4's answers to the 40 questions were practically idiosyncratic compared to the answers that other informants gave.

We can use this information to compute a competency score for each informant. To do this, run a factor analysis on the matrix of corrected matches. (*Anthropac* does all this automatically. You don't need to understand factor analysis to read the rest of this section. For an introduction to factor analysis, see chapter 21.) If the three conditions I've listed for the model have been met, then the first factor in the solution should be at least three times the size of the second factor. If it is, then this means that: (1) The first factor is *knowledge* about the domain (because agreement equals knowledge under conditions of the model); and (2) The individual factor scores are a measure of knowledge for each person who takes the test.

At the far right of table 8.4, we see that informants 2 and 3 have the highest factor scores (.61). They are also the students who got the highest of the four scores in the general knowledge test. You can use the consensus test on any group of informants, for any cultural domain. Triad tests, paired comparisons, ratings, and rankings all produce data that can be subjected to consensus analysis, as do true-false tests and multiple choice tests.

I want to stress that if you are doing general descriptive ethnography, and you're looking for all-around good informants, the cultural consensus method

is *not* a substitute for the time-honored way that ethnographers have always chosen key informants: luck, intuition, and hard work by both parties to achieve a working relationship based on trust. The cultural consensus method, though, is truly useful for finding highly competent people who can talk about well-defined areas of cultural knowledge. (For a detailed explanation of the math behind consensus analysis, see Weller 2004 and http://www.analytictech.com/borgatti/consensu.htm.)

Testing the Cultural Consensus Model

The cultural consensus model makes a lot of sense, but it may be a bit of a stretch to imagine that you can find the answer key to a test under certain conditions. You can test this. Get the results from any multiple-choice test in any class that has at least 40 students and run the consensus analysis available in *Anthropac*.

Correlate the first factor score for each student against the score that each student actually got on the test. If the exam was a good test of student's knowledge (that is, if the set of exam questions represents the cultural domain), you'll get a correlation of over .90. What that means is that the students who have the highest first-factor scores (knowledge scores) will mirror the professor's answer key for at least 90% of the items.

If you can retrieve an etically correct answer key, then you can apply the model (cautiously, of course, always cautiously) to tests of emic data, like people's ideas about who hangs out with whom in an organization or what people think are good ways to cure a cold, avoid getting AIDS, take care of a baby, etc.

The cultural consensus model is an important contribution to social science methods. It means that, under the conditions of the model (informants share a common culture and there is a cultural answer to each question; informants answer test questions independently of one another; the questions in the test come from a single cultural domain), you can build the answer key to a test from the matrix of agreements among informants. (For more about the consensus model, see Weller 2004. For more examples of consensus analysis, see Caulkins 2001, de Munck et al. 2002, Furlow 2003, Swora 2003, Harvey and Bird 2004, Jaskyte and Dressler 2004, and Miller et al. 2004.)

Selecting Domain-Specific Informants

Weller and Romney (1988), two of the developers of the cultural consensus model, have determined the number of informants you need to produce valid

and reliable data about particular cultural domains, given that the three conditions of the model are more-or-less met. (I say "more-or-less" because the model is very robust, which means that it produces very similar answers even when its conditions are more-or-less, not perfectly, met.)

Table 8.5 shows those numbers: Just 10 informants, with an average competence of .7 have a 99% probability of answering each question on a true-false

TABLE 8.5

Minimal Number of Informants Needed to Classify a Desired Proportion of Questions with a Specified Confidence Level for Different Levels of Cultural Competence

Proportion of questions	*Average level of cultural competence*				
	.5	*.6*	*.7*	*.8*	*.9*
.95 confidence level					
0.80	9	7	4	4	4
0.85	11	7	4	4	4
0.90	13	9	6	4	4
0.95	17	11	6	6	4
0.99	29	19	10	8	4
.99 confidence level					
0.80	15	10	5	4	4
0.85	15	10	7	5	4
0.90	21	12	7	5	4
0.95	23	14	9	7	4
0.99	>30	20	13	8	6

SOURCE: S. C. Weller and A. K. Romney, *Systematic Data Collection*, p. 77. © 1988. Reprinted by permission of Sage Publications.

test correctly, with a confidence level of .95. Only 13 informants, with a relatively low average competence of .5 are needed if you want a 90% probability of answering each question on a test correctly, with a confidence level of .95.

Weller and Romney also (1988) showed that you can use the simple Spearman-Brown Prophesy formula, available in many general statistical packages, as a proxy for the full consensus method (the one that involves doing a factor analysis on the informant-by-informant agreement matrix, and so on) when you have interval level data. Table 8.6 shows the results: If you interview 10 informants whose responses correlate .49, then the aggregate of their answers are likely to correlate .95 with the true answers.

Adam Kiš (2005) studied funerals in a village in Malawi. Funerals used to be events that brought everyone in the community together, but AIDS has changed that. There are now so many funerals that people have to decide

TABLE 8.6
Agreement among Individuals and Estimated Validity of Aggregating Their Responses for Different Samples

	Validity				
Agreement	*0.80*	*0.85*	*0.90*	*0.95*	*0.99*
0.16	10	14	22	49	257
0.25	5	8	13	28	148
0.36	3	5	8	17	87
0.49	2	3	4	10	51

SOURCE: S. C. Weller and A. K. Romney, *Systematic Data Collection*, p. 77. © 1988. Reprinted by permission of Sage Publications.

which ones to attend. Kiš developed a cultural domain test of "reasons to attend a funeral" and administered it to 30 informants. The results showed that there was, indeed, a consensus about this domain of funeral culture, so Kiš focused on the most knowledgeable informants, as determined by the consensus analysis, for his in-depth, ethnographic interviews.

Caution: If you use consensus analysis to find knowledgeable informants, watch out for the **shaman effect**. People who have very specialized knowledge about some field may be very different in their knowledge profile from people in the mainstream—that is, shamans. In fact, it is to the advantage of shamans everywhere, whether their knowledge is about curing illness or making money on the stock market, to protect that knowledge by keeping it maximally different from mainstream knowledge. The bottom line: Use consensus analysis to find the highly knowledgeable informants, but never pass up the chance to interview a shaman.

Lots more about consensus analysis in chapter 11 on cultural domain analysis.

Paying Informants

Finally, there's the issue of whether to pay informants, and if so, how much? 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. 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, they should be paid at the local rate for their time. And speaking of interviews. . . .