CHAPTER 6

Explicit Reports

Surveys, Interviews, and Tests

Learning Objectives:

- What are general properties and specific types of explicit reports?
- What are options for formatting explicit-report items?
- What are options for administering explicit reports, and what are some consequences of the different options?
- What are some ways to design and generate explicit-report instruments?
- What are characteristics of the U.S. census and the data it produces?
- What are basic limitations of explicit reports as a type of research data?

One of the most flexible and popular types of data collection in human geography is the explicit report. As we described in Chapter 3, explicit reports are measures of beliefs people have about all sorts of things, including themselves or other people, places or events, activities or objects. Explicit reports, including surveys, interviews, and tests, can request many different types of beliefs: behaviors, knowledge, opinions, attitudes, expectations, intentions, experiences, and demographic characteristics. The defining trait of such measures is their **explicitness**. People know they are providing information to a researcher when they are surveyed, for instance, and the responses they provide are based on opinions or beliefs they can consciously access, that is, of which they are aware. Often, explicit reports request responses that cannot readily be judged as being right or wrong; that is, the responses are personal opinions or preferences that cannot be compared to any objective standard of reality, although they can be characterized as common or unusual, and they can be related to other variables such as demographics. As we

Table 6.1 Major Types of Explicit Report Instruments

1. Surveys, questionnaires
2. Interviews
3. Sociometric ratings
4. Activity diaries, logs
5. Contingent valuation
6. Focus groups
7. Protocol analyses
8. Tests

described in Chapter 3, when the explicit responses *can* be assessed for correctness, and that is of major interest to the researcher, we call the reports tests. That is, tests are used to study knowledge rather than opinion.

Major types of explicit reports are listed in Table 6.1. The first is the **survey**, or questionnaire. Surveys require respondents to answer questions about their opinions, attitudes, or preferences. They can also ask questions about activities or demographics. Even though some of these have answers that are potentially true or false, that is not of primary interest to researchers—researchers typically accept them uncritically as true (we discuss skepticism about the veracity of responses at the end of the chapter). We use the term "survey" to refer to reports that are administered and responded to in written format. In contrast, **interviews** collect the same types of information as surveys but are administered and responded to orally.[1]

Three additional types of explicit reports are really subtypes of surveys. One is the **sociometric rating**. These are opinions or beliefs expressed by members of small groups, such as families or carpool groups, about each of the other members of the group. Such ratings are particularly valuable for studying small-group structures and dynamics. They have an interesting analytic property that results from all members rating, and being rated by, each of the other members. This allows a person's average ratings to be separated into a component due to how he or she rates others, a component due to how others rate him or her, and a component that is unique to the particular "rater-rated" dyad.

A second subtype of survey is the **activity diary** or log. Activity diaries require respondents to record what they do on a regular basis, typically on a daily basis. Geographers are often especially interested in *where* people carry out activities,

[1]Our distinction between surveys as written and interviews as oral is not universally held. For example, some authors refer to all explicit reports that do not focus on factually correct or incorrect information as surveys (or "self-report" measures), reserving the term "interview" for those surveys administered orally by a researcher, and the term "questionnaire" for those self-administered in writing. What is important is not so much what they are called, but that written and oral administration can produce somewhat different responses in different situations, and their costs certainly differ.

such as where they shop or which road they take. Transportation surveys are a good example. Activity diaries could be considered behavioral observations on one's self, but because they require people to explicitly report on their activities, we include them here rather than with the behavioral observations of Chapter 5. Similarly, activity diaries as explicit reports should not be confused with the everyday, common use of the term "diary." As the term is commonly used, diaries are confidential records or personal letters that people write to themselves to record their experiences, activities, thoughts, and feelings. Such diaries are nonreactive measures, of course, because their authors do not expect that researchers will use their diaries as data. We cited traditional diaries as a source of archival data in Chapter 5.

A third subtype of survey is **contingent valuation**. This type of report requires people to rate or rank how much they value something, typically something that may have great subjective value but is very difficult to assign value to "objectively." Contingent valuation is commonly used in studies of how people value particular landscapes or environmental actions, for example, preserving open space or maintaining quiet parks. It has been applied in other contexts too, such as studies of mass transit use. Ratings of value are expressed directly in dollars or indirectly in terms of units of something else that the respondent values, such as hours of time or miles from family members.

Two more types of explicit reports are really subtypes of interviews. **Focus groups** are unstructured (discussed below) interviews carried out with small groups of respondents, with as few as two or three members, or as many as 30 or more (7–20 is a common range). Typically, focus groups discuss a particular topic and are led by a facilitator or moderator. The most notable characteristic of focus groups is that they involve a simultaneous interview with a group of people, so various **group dynamics** develop. Of course, members make comments that jog other members' memories. But the comments also influence what other members think or are willing to say; nonverbal cues do this too. For both conscious and unconscious reasons, people will often change what they say and how they say it in order to make impressions on others they know are observing. These various social effects are amply documented in a vast literature in social psychology and sociology on persuasion, conformity, and self-presentation. Furthermore, the outcome of a focus group depends on the particular people who make up the group; for example, outspoken group members can monopolize a focus-group discussion. Transcripts or notes of focus groups are usually not carefully coded (Chapter 5) or analyzed quantitatively; they should not be unless extensive steps are taken to deal with their limitations. In most cases, focus-group results are only very useful during exploratory phases of research.

A second subtype of interviews is **protocol analysis**. This is an open-ended (discussed below) interview in which people "think aloud" about the contents of their conscious mind while reasoning about some problem or issue. It attempts to identify what people are thinking about during some mental task, including what they are looking at, listening to, imagining, inferring, or retrieving from long-term memory. For example, a researcher could ask people to talk about their thoughts while navigating a geographic information database system, a study that could help make the system's design more user-friendly. Protocol analysis is not about general

or average thoughts, or about justifications or rationalizations. Done properly, it is about the *products* of thought at different stages, not the *processes* of thought. As we consider at the end of the chapter, mental processes are typically unavailable to conscious access, and in fact, so are many mental products. Many uses of protocol analysis ask people to talk about (or point, draw, and so on) what they are thinking or attending to while they reason; this is called "concurrent" protocol analysis. Alternatively, people may talk about their thinking *after* they finish reasoning, perhaps because the talking would interfere with the primary task; this is called "retrospective" protocol analysis and may be relatively immediate or delayed. Whether the protocol is concurrent or retrospective, it is typically quite valuable for the researcher to perform a detailed analysis of the primary task the person is working on as an important way to help interpret the protocol.

Finally, as we have defined them, **tests** require participants to respond to questions that can be assessed as right or wrong. Researchers administering tests are specifically interested in the factual correctness or incorrectness of responses to their tests. A geographer might use tests, for example, to study a person's knowledge of population demographics or economic development (of course, if the geographer is using the test to evaluate your performance in a class, we call it *teaching* rather than research). Tests are typically scored for accuracy, including number or size of errors, in order to generate the data. Sometimes people are timed while taking tests, and their response times are recorded to produce data. Response times can be theoretically interesting, especially when used comparatively, because they reflect how long it takes someone to remember or reason about something. For example, if a GIS user takes longer to solve a problem with one interface than with another, that is evidence for the relative superiority of the faster interface, other things being equal.

Format of Explicit Reports

All types of explicit reports ask people to respond to questions, statements, pictures, or other stimuli. A specific survey, interview, or test is called an **instrument**; the specific questions or statements that make it up are called **items**. We have distinguished between explicit reports expressed in written versus oral form, but in both cases, reports are most often expressed verbally, in words. However, responses to explicit reports can be expressed in numbers, gestures (such as pointing), graphics, or manipulable objects. In the case of manipulable objects, respondents can be asked to mark lines, draw pictures, sketch on maps, or construct physical models. In this section, we discuss some alternatives for the formats of these items.

Perhaps the most important aspect of item format is the distinction between closed-ended and open-ended items, one important way that many people distinguish between quantitative and qualitative research methods (as we discussed in Chapter 3). **Closed-ended items** provide a finite number of specific response options for respondents to choose as answers, typically a modest number. They provide responses in a fixed-response format. **Open-ended items** do not provide specific response options for respondents. They provide responses in an unstructured

or free-response format. As compared to closed-ended items, open-ended items allow responses that were not anticipated by the researcher. Also, open-ended items allow responses of any length, assuming enough time or space is allowed, and responses that are meaningful to respondents in their own words. However, as we explained in Chapter 5, open-ended responses are not data in themselves—they are records of (typically verbal) behavior. If they are to serve as anything more than impressionistic suggestions, they must be coded (we explained how to code in Chapter 5). Interestingly, some research has shown that closed-ended items often prompt respondents to choose response alternatives that they would not spontaneously make to open-ended items.[2]

Another distinction concerns the way the items are formatted for different respondents. **Standardized items** are presented in a predetermined and consistent format to all respondents. This consistency may be in the wording of the items, their response formats, the order of items, how they are administered, or any combination thereof. In one or more of these ways, **nonstandardized items** are not consistent across respondents. The most common examples of nonstandardized items are explicit reports that contain **follow-up questions**, the specific wording of which depends on responses to earlier items. Follow-up questions are of two types. Some follow-up questions have a so-called **branching format** in which questions vary "automatically" as a function of responses to earlier questions. For example, an item can be presented as appropriate depending on the answer a respondent gives to a previous demographic question, such as ethnicity or income; questions about retirement might follow only if people give their age as over 60. The second type of follow-up question is even more nonstandardized. These are **free-format** follow-up questions that attempt to clarify and dig deeper into the meaning of responses. Their specific wording is determined in real time by an interviewer who reacts spontaneously to a previous answer.

Table 6.2 lists common types of closed-ended items. Closed-ended items should always provide exhaustive alternatives to respondents and should usually provide mutually exclusive alternatives too, so that the respondent can comfortably pick one and only one alternative. In some contexts, however, it may be perfectly reasonable to allow respondents to pick more than one alternative; this is typically true with adjective checklists, for example. Whenever a respondent might plausibly want to answer something you have not provided as an alternative, include an "other ___" response; this makes a closed-ended item somewhat open-ended. Although some researchers prefer to force respondents to express a belief about something, we believe respondents should be allowed to choose "don't know" or "no opinion" in most cases. When analyzing these alternatives, be careful that you separate "don't know" or "no opinion" responses from the rest as appropriate, such as when you calculate proportions or averages.

---

[2]It is not clear whether this means that closed-ended alternatives remind people of things they really do believe, or that they put ideas into people's heads they would not otherwise have.

Table 6.2 Major Types of Closed-Ended Items

1. **Rating scales**. Respondents provide a number or mark a line to indicate the amount or extent of something, including the degree of belief they have in something
2. **Forced-choice alternatives**. Respondents pick one alternative from a list of choices, such as "yes-no," "true-false," or "multiple choice."
3. **Ranking of alternatives**. Respondents put two or more alternatives in order from most to least, or least to most.
4. **Adjective or activity checklist**. Respondents mark each alternative that applies from a list of adjectives or activities.
5. **Paired or triadic comparisons.** Respondents specifically compare entities, two in the case of paired comparisons, and three in the case of triadic comparisons. These comparisons can be ordinal ("Which two of the three neighborhoods are most similar?") or metric ("What is the distance from the courthouse to the visitor center?")
6. **Sorting task**. Respondents sort phrases, pictures, or other stimuli into groups based, typically, on similarity.

Closed-ended items like rating scales occasionally produce patterns of responses that suggest a respondent was not answering sincerely. These are called **response sets.** They come in various forms. **Social desirability** is the tendency for respondents to give answers they think the researcher, or some imagined group of peers or superiors, want to hear. This may be reflected in a set of answers that are "just too good to be true," which we admit may not be easy to detect. Response sets are often reflected in forms of consistent responding, such as yea-saying, nay-saying, extremism, or moderatism. For example, a respondent might only use the highest value on a rating scale. To deal with this, it is sometimes recommended to alternate the "sense" of items so that a response to one side would not always reflect agreement with the same view. This probably does induce respondents not to respond reflexively, but it also has the real potential to confuse them. As another example, someone responding to a survey of "environmentally responsible behaviors" might make dubious claims about always recycling (every candy wrapper), never leaving lights on, sharing rides everywhere, and taking showers in cool water. Including instructions that urge respondents to use several values on the scale could help here. In any event, you have every right (even a *responsibility* of good data analysis) to remove participants from the data set when you detect response sets. However, this must be based on some rule that is objectively describable other than that someone gave responses counteracting your hypothesis; you will describe this rule in the Results section of the research manuscript (Chapter 13). Of course, insincere responses are not always possible to spot as some pattern that sticks out. Such responses are hopefully (and we think probably) rare, but they undoubtedly happen. When they do, they typically add "noise" to the data, which lowers reliability (Chapter 11).

Rating Scales

Let's consider rating scales in a little more detail. As we define them in Table 6.2, rating scales ask respondents to provide a number or mark a line to indicate the amount or extent of something, including the degree of belief they have in something. Several types of rating scales are shown in Table 6.3. We use the term **generic rating scale** to refer to any rating scale that does not have one of the more specific formats and purposes of the other types. The **semantic differential** requires respondents to rate a set of entities (objects, events, places) on each of several adjectives. Semantic differential scales are designed to measure attributes of connotative, rather than denotative, meaning—not what an entity literally means but what it suggests or implies. **Likert scales** require respondents to express their degree of agreement or disagreement with a series of statements. **Paired comparison ratings** require respondents to compare entities (as described in Table 6.2) with the use of a quantitative rating scale.

There are a couple of general design issues that apply to all types of rating scales. First is the question of how many scale options to provide respondents. In most situations, the answer is five to nine options, preferably toward the low side when orally administering an instrument. The rationale behind recommending five to nine options is that fewer options provide less measurement resolution than people can validly make, while more options provide too much resolution. Although one may be tempted, for example, to ask respondents to rate on a 100-point scale (people are used to "100-point thinking" in our culture), evidence suggests that people cannot validly make discriminations at that fine a level.[3] Although respondents might claim to distinguish a preference of 70 from one of 75, never mind 72 or 73, they probably cannot do so in a way that validly reflects their feelings. It would be spurious precision, as we defined in Chapter 2. When working with young children, one should use a scale with two to five values; a charming series of happy, neutral, and unhappy faces is sometimes used. Whatever the specific number of scale options one uses, it is best to use an odd number in order to leave room for a midpoint of neutrality, when that is appropriate. And as for other closed-ended items, you might or might not find it a good idea to offer respondents the opportunity to choose "don't know" or "no opinion" as response options (these are not shown in Table 6.3).

Finally, we want to come back to the issue we touched on in Chapter 2 concerning the measurement level of data obtained from rating scales. As we noted in Chapter 2, several authors of statistical and methods textbooks in geography (and in other disciplines) advise treating rating-scale data as ordinal. That is, the interval between "3" and "4" should not be assumed to equal the interval between "4" and "5." Following this advice has the fairly substantial implication that only nonparametric ordinal statistics (see Chapter 9) should be applied to rating-scale data:

---

[3]These limits on human discrimination are discussed in a famous paper that also deals with limits to working memory capacity: Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Table 6.3 Types of Rating Scales

Generic Rating Scale

For each place, circle a number from 1 to 9 to rate how much you would like to take a vacation there in the next year:

not at all	1	2	3	4	5	6	7	8	9	very much

_____ Disney World, Orlando, Florida

_____ Yellowstone National Park, Wyoming

_____ Las Vegas, Nevada . . .

Semantic Differential

Rate each city on each scale by marking a number from 1 to 7:

big	1	2	3	4	5	6	7	small
hot	1	2	3	4	5	6	7	cold
safe	1	2	3	4	5	6	7	dangerous . . .

Likert Scale

Rate your degree of agreement or disagreement with each statement by circling one phrase:

strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
1	2	3	4	5

"Individual people are more responsible for energy conservation than is the government."

"The U.S. should reduce its dependence on foreign oil."

"The federal government should increase funding to develop efficient solar energy."

Paired Comparison Ratings

For each pair of cities, rate how similar the two cities are to each other by marking a number from 1 to 7:

completely different	1	2	3	4	5	6	7	exactly the same

_____ London, U.K. – Bangkok, Thailand

_____ Lima, Peru – Moscow, Russia

_____ Cairo, Egypt – New York, U.S.A. . . .

no means, no standard deviations, no analysis of variance (ANOVA), no Pearson correlations or least-squares regressions (it also has implications for graphing and mapping—see Chapter 10). Contrary to this, several other textbook authors, mostly outside of geography, consider it acceptable to treat rating-scale data as metric. In fact, rating-scale data have been treated as interval level in thousands of published studies by top researchers in psychology, sociology, and other fields (rarely would it constitute ratio-level data, as "0" on a liking scale from −3 to +3 is an indifferent preference, not "no" preference at all). These researchers consider this treatment to be both justifiable, accurately reflecting the underlying nature of the attitudinal and preference constructs being measured, and functional, allowing the most powerful and flexible statistical analyses possible.

We endorse the view that rating-scale data generally can be treated at the interval level, particularly when the scale is an attempt to measure an underlying construct that can reasonably be considered continuous, such as preference or degree of confidence. We certainly agree that people do not treat equally spaced scale values as *exactly* equally spaced differences in the property being measured. However, we also believe, and some evidence supports, that people do treat equally spaced scale values as *approximately* equally spaced differences, especially when the scale presents visually and/or numerically equally spaced values to the rater (for that reason, we do not recommend the use of wordy verbal labels for each scale value that might not be interpreted as evenly spaced concepts). That is, we believe that rating scales as people understand and use them are not precise interval scales, but that they are much more than merely ordinal. If they were just ordinal, the underlying difference between "3" and "4" could as easily be 10 times the difference between "4" and "5" as it could be one-tenth the difference—and this is untenable. Instead, we see intervals in rating scales as being understood and used by people somewhat vaguely and with some variation across times, individuals, and even positions in the scale. This conception of a rating scale would be captured visually by an animation in which intervals between scale values were a little fuzzy and quivering. Given this conception, we consider it justified treating rating-scale data as interval rather than ordinal, accepting that we thereby introduce much greater power and flexibility into our research at the cost of a little conceptual or empirical error. If you, or your advisor or editor, don't accept this position, feel free to treat the rating scale data as ordinal. At worst, you might fail to uncover some interesting patterns in your data.

The Administration of Explicit Reports

Besides their format, there are a variety of other options about how explicit reports are administered. They may be self-administered or researcher-administered (of course, people do not generally *interview* themselves). They may be administered individually or in groups. The medium of administration could be in person (face-to-face), through the mail, over the telephone, or on the Internet (more below on Internet research). Interviews are often done with the help of computer programs that display the questions and accept typed-in answers. Also, interviews may be recorded, either audio or video. A variety of considerations determine the best way to administer explicit reports in a given research situation:

1. *Cost.* The cost of doing research with explicit reports, whether expressed in terms of money, time, or effort, can vary greatly as a function of the way they are administered. A case in point is oral interviewing versus written surveying. Interviewing is usually much more costly; even if interviewers are not paid, they must be selected, trained, and supervised. But the cost of sending out mail surveys, including the cost of postage, is not trivial. Internet surveys have become an attractive alternative because of their low cost.

2. ***Number and nature of items.*** Usually, surveys can contain more items, and more complex items, than interviews. This is because of the extra demands placed on memory by listening and responding orally, as compared to reading and writing. Complex items or complex response formats are difficult or impossible to administer orally, although there are techniques for orally administering response options more effectively. Even more than complexity, the various administration options can differ greatly in their effectiveness at getting honest and forthcoming answers to sensitive questions. Such questions include not only such obviously personal matters as sexual behavior and substance use, but personal finances, political beliefs, and environmentally responsible behaviors. The degree of anonymity or confidentiality a respondent perceives varies quite a bit across administration options, and that definitely affects the honesty of responses (anonymity and confidentiality are relevant to the ethical treatment of human research subjects, as we discuss in Chapter 14). Again, there are various techniques that can be used to increase respondents' forthcomingness, including clever ways to ask questions that don't require respondents to provide a personally identifiable answer that could even possibly be linked to them.

3. ***Response rate.*** Response rate is an important consideration in choosing an administration method. More people throw mailed surveys away, considering them junk mail, than refuse to talk on the phone, although frustration with incessant "sales calls" has reduced the response rate to phone solicitation over the last decade. People apparently ignore e-mail solicitations at an even higher rate than they do paper mail solicitations. In-person requests for participation are the most effective in terms of response rate, but of course, they are costly.

4. ***Potential for follow-up.*** Administration options differ a great deal with respect to the possibility they allow for follow-up questions that depend on the nature of responses to earlier items. This can be among the most important considerations in choosing among options. Here we are talking mostly about free-format follow-up questions that attempt to clarify and dig deeper into the meaning of responses. That can be done effectively only with oral administration, whether in person, on the telephone, or during Internet exchanges (e-mails can be sent to follow up on responses). In contrast, branching-format questions that "automatically" pose follow-up questions tailored to particular responses to earlier questions are possible with any and all administration options. Until artificial intelligence becomes much more advanced, however, it is generally impossible for a computer program

to interpret the meaning of a response and then ask a follow-up in order to clarify part of that meaning. A sentient human being who is thinking about the response as it is given is required.

5. ***Nature of respondents.*** This is a fairly obvious, common-sense consideration. But common sense is sometimes surprisingly rare, and in any event, no consideration is obvious if one fails to think of it. Different types of respondents differ in their ability to handle different administration methods, different types of questions, and so on. Respondents differ in ways that have implications for administration options. Relevant ways include age, sex, formal or informal education, social status or position, language spoken, disabilities (sensory, motor, cognitive), and so on. Different cultural and subcultural groups of people have been exposed to different norms that can have large effects on the viability of different administration options. Depending on the sex of the respondent, an interviewer of a particular sex can be quite inappropriate in some cultures. Revealing some kinds of information to a researcher who is not actually present may be strange to the point of not working.

6. ***Possible interviewer artifacts.*** In Chapter 11, we discuss potential threats to the validity of research that might arise because of the appearance, demographic characteristics, or personal style of researchers. These "researcher artifacts" are an issue only when the researcher is in the presence of the respondent when he or she responds (not quite—the respondents' expectation that different kinds of people will look at their responses could have an impact). For example, male respondents may not answer questions about their "sense of direction" in the same way when speaking to female interviewers as they do to male interviewers; female respondents may not either. The wording of items may reflect bias on the part of the researcher, which may in turn bias responses; the nonverbal and "paraverbal" cues (body language, intonation) used by an interviewer when orally administering items, including follow-up questions, may bias or lead responses. In addition, some methods of administering or scoring explicit reports allow or require the researcher to interpret what respondents said or did in order to turn it into recorded data. This can obviously open the door to the potential of biased interpretation on the part of the researcher. We discuss these researcher artifacts more in Chapter 11, including some approaches to minimizing them.

Using the Internet to Collect Explicit Reports

Within the last decade, the possibility of using the Internet, including the World-Wide Web, to carry out research with explicit reports has become a reality. The major benefit of this approach, whether e-mailing surveys or having people respond to a Web form, is its efficiency and low cost. This benefit is potentially huge. Researchers have recently been using the Web to solicit and record responses by thousands or tens of thousands of people, at extremely low costs. What's more, data entered on Web surveys, including forced-choice options and open-ended text, are

automatically stored in digital form, and to varying degrees, can be analyzed automatically as well. Web surveys and tests can be automatically modified as they are being administered to take account of earlier answers—the branching-format follow-up questions we discussed above.

But Internet research is not without risks to the quality of the data. There has been considerable concern that samples of respondents obtained over the Internet, often people who choose to go to a particular Web site and fill out its survey, are unusual in some ways and do not represent some larger population very accurately. In other words, there is concern about the presence of a sampling bias in Internet research. In fact, empirical research[4] has demonstrated fairly convincingly that Internet samples are not perfectly representative of the general population. However, they are better than college student samples in this respect, being more heterogeneous on most variables, including sex, race and ethnicity, socioeconomic status, and age. In particular, Internet samples are superior with respect to regional diversity. In addition, assessments of personality and psychological adjustment suggest that Internet samples these days are not filled with "geeky antisocial nerds." As much as Internet users were once an unusual subset of the population, finding people who never use the Internet will become increasingly difficult in the future.

There are other potential difficulties with Internet research. Perhaps respondents participate repeatedly, either because they don't know that it is inappropriate or they are protégés of the Daley school of survey research ("respond early and often"). However, research suggests that this does not occur that commonly, and in any case, there are techniques for identifying repeat responders and omitting them. Some research indicates that participants in Internet research drop out at higher rates, but this is not much of a problem, given the high numbers of people such research can contact in the first place and the fact that nonparticipation bias is apparently no greater with this type of research than with more traditional types. Researchers are sometimes concerned that Internet respondents will take studies less seriously and provide more frivolous responses, but most research indicates that the results of Internet studies are similar to traditional studies on comparable topics. In addition, there are potential ethical difficulties (Chapter 14) with Internet research that are getting quite a bit of attention, including the potential for loss of privacy and the fact that Internet researchers may not always be able to ensure that respondents can provide genuine informed consent in order to participate.

Finally, the Internet provides quite a bit more to researchers than just an efficient way to administer surveys and tests. The Internet in its various guises has led to the emergence of a host of new social and behavioral phenomena that are of interest in their own right to researchers, including geographers. Such new phenomena are the subject of ongoing and future research, and include the diffusion of innovations via

[4]A good example is Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

the Internet, social interaction in digital worlds, and the emergence and structure of online communities.

Designing and Generating Explicit Instruments

Having considered various structures for report items, we now turn to the question of how to create the items themselves. How do we get ideas for items? The answer is pretty much the same as the answer we gave in Chapter 2 as to how to generate research ideas—any way that works! Intuition and prior knowledge, whether educated or naive, is a basic starting point. Existing literature will provide many ideas, including specific items and even entire instruments that already exist. Even if you don't find an instrument in the literature you can use in your research, you should make sure that whatever you come up with to use is compatible with whatever wisdom is to be found in that literature. For example, make sure your items deal with the concepts the literature says are important in a particular domain. A good source of ideas for generating items is to conduct one or more open-ended, unstructured interviews with single respondents or focus groups.

However you generate items, the way you construct them is extremely important, because that will greatly influence how respondents understand them and come up with responses to them. That has major implications for the reliability of our instruments, their construct validity, or both (see Chapter 11). Avoid confusing, biased, and ambiguous wording by using clear and unambiguous language, understood consistently by all respondents. For example, if you ask respondents to state whether their "income level is low, medium, or high," you have the problem that different people interpret low or high incomes very differently. One should steer clear of items that involve double negatives; a classic example is the multiple-choice question that inspires respondents to pull out their hair because it asks "which is not true about. . .?" and then provides "none of the above" as an option. Items need to be **unidimensional**—they need to ask only one thing. An item that asks "do you favor preserving farmland and forests?" won't do, because a person could easily favor one landcover differently than the other. Avoid biased and emotionally charged wording, as such items can be leading or create unfortunate reactions in respondents, like quitting the data collection session. Who says no to this: "Do you favor preserving pristine and beautiful wilderness areas?" Finally, design the visual appearance of surveys so they are easy to understand and use (apply graphical principles like those discussed in Chapter 10), and so they communicate a serious attitude on the part of you, the researcher.

What about the length of explicit report instruments? They obviously must be long enough so you can obtain the information you need, but they should not be longer than that. This is harder than it sounds; many researchers, especially when they are inexperienced, readily fall into the attitude that "well, while we're at it, we might as well ask them that too." Unnecessary questions create an unwieldy and unfocused instrument that can bore or overtax respondents. Excessive length can also decrease the response rate and lower the thoughtfulness of responses.

And given that administering, coding, and analyzing explicit reports is usually very laborious and cognitively challenging, brevity and conciseness are surely blessings.

The order of items in an explicit-report instrument is often important. Sometimes items must follow other items because they depend on information in those earlier items or in the respondent's answer to earlier items. But even when item order is not dictated, it may influence responses because of the context created by what previous questions induced a person to think about or feel, because of fatigue, and so on. Respondents are typically much more comfortable with the format of rating scales by the end of an instrument than on the first item, for example. These are called **order effects**. They come in many potential forms, including some that are due to the absolute position of an item within a sequence of items, and others that are due to the specific context created by the immediately preceding item or items (in which case they may be called **context effects**). There are several approaches to dealing with order effects. One can randomize the order of items for each respondent, which is relatively easy to do with computer-administered instruments, or at least use several different random orders; of course, which respondent gets which order must be decided randomly. Alternatively, if one has a small number of items, one can give each possible order to equal numbers of respondents. When this approach, known as **counterbalancing**, is feasible, it is optimal because it perfectly balances all possible order effects.[3] Sometimes items may group logically into subsets; the item order within each subset may be fixed for all respondents, but the order of the subsets may be counterbalanced or randomized.

We conclude this section by offering a strategy for creating explicit report instruments. We emphasize that creating instruments is very hard work and really cannot be done well on the first try, even by experienced "research geniuses." We stress this point because we have seen so many inexperienced researchers jump right into research with explicit reports, generating items, designing the instrument, and beginning final data collection within a period of one or two weeks. That just doesn't work, but we sure wish we had a euro for every time an inexperienced researcher proceeded in that fashion, only to bemoan later some ambiguity or mistake that prevented them from drawing the conclusions or answering the questions they set out to. We recommend you follow a cycle that we call "GPM . . . PMT." That stands for "generate, pretest, modify . . . pretest, modify, test." In other words, first use any combination of the methods we discussed above to generate items you think will tap into your constructs of interest. Then pretest your first draft; such a preliminary study designed to practice and evaluate any component of your primary data collection is called a **pilot study**. Do this by showing it to your friends, colleagues, or office mates. Administer it to a small sample of participants from a research pool or class of students. At some point, you should definitely pretest your instrument with a small sample of the type of respondents you will eventually use for real data collection.

[3]As the number of items grows, the number of possible orders can quickly become too large to do this. There are elaborate strategies for generating less than all possible orders, such as the *Latin square* technique, that nevertheless deal more systematically with various potential order effects than does complete randomization.

To pretest, don't just administer your instrument, but ask your pilot respondents how they interpret items, instructions, and so on, and whether they are confused about any part of the instrument or its administration. In other words, carry out a pretest that is open-ended and unstructured. The next step is to make modifications to the instrument in response to feedback you get from the pretest. Then pretest it and modify it again. You should sometimes repeat this sequence of pretesting and modifying several times, especially with large, complex, or expensive studies. That's why we put the ellipsis in GPM . . . PMT. When you feel that your instrument is clear, adequately focused, and of the proper length—but only then—you can proceed with the actual testing phase. By the way, make sure to pretest the instrument with whoever is going to administer it, too (research assistants, interviewers).

The Census: An Important Secondary Source of Explicit Report Data for Geographers

A **census** is a count of the number of people in a country and an assessment of their characteristics, such as family structure, economic activities, and so on. They are carried out by national governments within a fairly short time period, so that the population does not change too much from the start to the end of data collection. Censuses provide an important source of explicit-report data for geographers. It is a secondary source, as census data are not collected for the specific purpose of a particular research project. The answers to census questions vary both spatially and temporally. Many governments find the answers to these questions to be sufficiently important to expend substantial resources in an attempt to answer them. Historically, governments have wanted to answer some of these questions for taxation and military purposes. With the advent of democratic governments, censuses have been required to fairly apportion representatives to administrative regions, for example, U.S. states. Increasingly, census data are used for planning, public health, and many scientific purposes.

Governments of ancient Egypt, Babylonia, China, India, and Rome conducted censuses at irregular time intervals; systematic and regular census enumerations did not really start until the middle of the 18th century. The U.S. Constitution (Article 1) mandates a census of the population in order to apportion taxes and congressional representatives. Although we focus on the *U.S.* census in this chapter, many other countries conduct censuses of various kinds; in fact, the United Nations has encouraged several countries to conduct regular censuses. The United States has conducted a census of its population every 10 years since 1790. The actual questions used in the census are approved by Congress for each enumeration; consequently, the questions asked have varied over the years. Substantial changes have occurred for questions associated with numbers of slaves owned, fertility, foreign-born status, mental and physical capability of household members, income, housing characteristics, and racial and ethnic identity. These changes introduce some difficulties for longitudinal studies (Chapter 7). At the same time, they are also a fascinating documentation of changing social and political values over the course of U.S. history.

The U.S. census is designed as a complete enumeration of the population; this precludes the use of statistical sampling for some critical population information. The question of complete enumeration versus statistical sampling has recently become a significant political issue.[6] From 1790 to 1930, all census questions were asked of all appropriate persons. Over the years, Congress added questions to the census, the country's population grew, and conducting a census with so many questions became logistically and financially difficult. Consequently, in 1940 the **Census Bureau** began to administer a small set of basic questions, often referred to as the **short form**, to the entire population and a much larger set of questions, the **long form**, to a fraction of the population (about one in six). The short-form items concern basic demographic and housing information: name, age, sex, race, year of birth, marital status, relationship to head of household, Spanish/Hispanic/Latino origin, and a few questions regarding housing characteristics. (It is interesting that for the first time, the 2000 census allowed individuals to indicate more than one racial category; in fact, fewer than 3% of respondents did.) It is important to us as geographers that demographic and housing characteristics are spatially located ("geo-referenced" in Chapter 12). The fact that the surveys are sent to specific addresses, not people, means that spatial location is implicitly measured.

The long form in the 2000 census had 26 additional population questions and 20 additional housing questions. Notable among these were questions about income, employment, migration, commuting and transportation behavior, education, birth location, military service, cost and financing of housing, and so on. Administering the long form presents some methodological problems, particularly **nonresponse bias**, discussed in Chapter 8 as "nonparticipation bias." The Census Bureau goes to great effort to ensure that long forms are filled out, including employing "nonresponse enumerators" who actually visit the addresses of forms that were not returned. Because of these difficulties and the coarse temporal resolution of the census (once a decade), the Census Bureau implemented a new continuous census in 1996: the **American Community Survey** (**ACS**). The ACS measures approximately 3 million people a year, about 1% of the entire population, with a survey that is essentially the long form. It is anticipated that the 2010 census will abandon the use of the long form altogether and send the short form exclusively. The ACS will replace the information obtained by the long form and increase its temporal resolution.

---

[6]The so-called **sampling debate** concerns this question of complete enumeration versus sampling and statistical estimation. The U.S. constitution calls for a "census," which implies a complete count. Nonetheless, as we discuss in this chapter, it is well established that some people are undercounted, particularly those who are poor, homeless, and so on. In the last couple of decades, several social scientists and statisticians have recognized that high-quality sampling and statistical estimation would very likely improve the accuracy of counts of underrepresented groups, and for much less money. Aside from the merits of the two arguments, the question of a complete count versus sampling and estimation has clear political implications insofar as liberals and conservatives believe they would on average benefit or suffer, respectively, from more accurate counts of groups such as the urban poor. See Wright, T. (1998). Sampling and Census 2●●●: The concepts. *American Scientist, 86*, 245–253.

Another important question regarding a census is "Who should be counted?" One approach is referred to as a **de facto census** of the population. A de facto census attempts to count who was in a given area at a given time. One way to think of this is to imagine dropping a giant cage over a particular area. Count and gather data from everyone that you "captured," and you have a de facto census of that area. A de facto census is not concerned whether people are tourists, students, illegal immigrants, and the like, only whether they are there. In contrast, a **de jure census** attempts to count people who legally belong to a particular area; that is, it does not count tourists, illegal immigrants, and so on. The U.S. census is neither purely de facto nor de jure; it uses a more vague definition of who it is attempting to enumerate—people at their "usual residence." This approach to the census is considered appropriate because of the fundamental and constitutionally mandated purpose of the U.S. census to demarcate congressional districts. This leads to some interesting issues for geographers who use the census as a source of research data. For example, in Las Vegas, Nevada, the census does not count the continuous and extremely large population of tourists. In the New England states, particularly Boston, the census may or may not count the students from around the country and world who attend the many colleges and universities there. Residents of prisons, who are not allowed to vote, are nonetheless counted (and rather accurately!) by the census. People who have immigrated into the U.S. illegally similarly do not vote but are counted (less accurately, to be sure). Likewise, some areas of the country have a much higher proportion of people under 18 years of age who are counted but do not vote. Congressional districts are drawn to contain equal populations of people counted in the census, regardless of whether or not they are allowed to vote. Thus, the question of "Who is counted?" raises the additional questions of "Who is not counted that should be?" and "Who may have been counted more than once?"

It is generally accepted that the U.S. census suffers from some degree of an **undercount** of the population that it is attempting to enumerate; this is a threat to the "construct validity" we define in Chapter 11. Actually, the net undercount results from both a larger undercount and an **overcount** that is somewhat smaller. It is likely that people who have immigrated illegally, people without permanent residences, and people of lower socioeconomic status living in inner cities are undercounted; one piece of evidence supporting the idea that the latter are undercounted is a comparison of military conscription data to census counts. An overcount results when students are counted twice, once by their parents at their "permanent" home location and once at their college or university residence. "Snowbirds" or other people who seasonally travel to another region of the country may be counted more than once. The Census Bureau goes to great effort to prevent both overcounting and undercounting; nonetheless, it is recognized that there is a measurable amount of both. It is also likely that the Census Bureau is making progress in diminishing the magnitude of the undercount. It's kind of intriguing to think about how the Census Bureau can know what it didn't measure. We won't discuss this issue in detail here, other than to note that measuring **coverage error** (overcount and undercount) is mostly accomplished by methods known as "demographic analysis" and "dual-system estimation."

What Do U.S. Census Data Look Like and How Can You Obtain Some?

The U.S. census is a substantial and sincere attempt to characterize the entire U.S. population demographically, geographically, and economically. The 2000 census counted over 281 million people in over 100 million households. In 1790, the census was likely recorded with feather pens on crude paper. The year 1950 marked the first use of digital computers to process and record census information. Since 1980, census data have been stored in geographic information systems (GISs, see Chapter 12). Initially they were stored with **DIME files** (Dual Independent Map Encoding files); now they are stored with **TIGER files** (Topologically Integrated Geographic Encoding and Referencing files). Storing census data in a GIS enables easier access, analysis, and mapping and visualization of the information. Privacy concerns are an important issue for the Census Bureau, and the actual completed survey forms are not made available to the public for 70 years. Also, census data are aggregated to spatial units designed to ensure the privacy of individual respondents. U.S. census data can be obtained from numerous public and private sources for a surprisingly wide range of costs, from free to rather expensive. A good start is the Web site of the Census Bureau at http://www.census.gov.

Census data are provided at a range of spatial resolutions. The spatial hierarchy of major census units is as follows (see Figures 6.1 and 6.2): entire United States; census regions (four); census divisions (nine); states (50 and the District of Columbia, plus nonstate U.S. territories like Puerto Rico and Guam); counties (basic administrative and legal subdivisions of states); county subdivisions (such as minor civil divisions, towns, townships); census places (incorporated cities and unincorporated census-designated places); **census tracts**[7] (small, relatively permanent subdivisions of a county, typically containing about 5,000 people); **census block groups** (clusters of blocks within census tracts); and **census blocks** (the smallest geographic unit of the census, typically bounded on all sides by streets, railroad tracks, streams, lakes, and so on). The boundaries of these units change somewhat from census to census. When conducting longitudinal studies, you typically find that the smaller your geographic unit of analysis, the more likely its boundaries will be different from one year's census to another; that is, blocks are very likely to be different, regions are very unlikely, and so on.

In addition, the census bureau classifies urban areas into a fairly confusing alphabet soup of acronyms. This is largely because cities are increasingly difficult to define and delimit precisely; their "commutersheds" have become so large that a hierarchy of categories is required to characterize the urban-rural distinction in any realistic way. In 1950, the Census Bureau used county-level data to come up with the concept and definition of a **standard metropolitan statistical area** (SMSA) as counties with a core city of at least 50,000 residents and a population density of at least 1,000 persons

[7]Census tracts are the smallest units at which long-form data are made available, for privacy reasons.

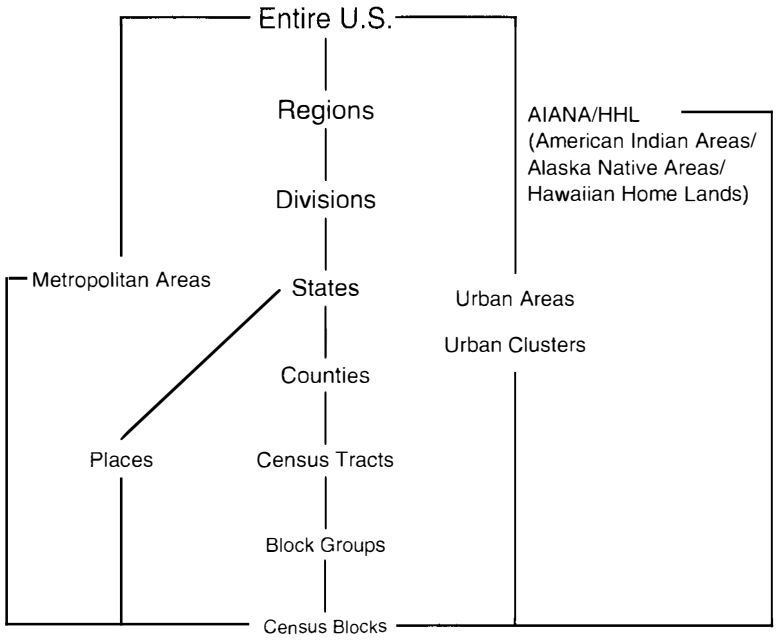

Figure 6.1 Some major geographical units used by the 2000 U.S. census (adapted from the U.S. Census Bureau Web site: http://www.census.gov)

per square mile. The SMSA was supposed to be "an area containing a recognized population nucleus and adjacent communities that have a high degree of integration with that nucleus." Subsequently, the Bureau created MSAs (metropolitan statistical areas), CMSAs (consolidated metropolitan statistical areas), PMSAs (primary metropolitan statistical areas), and CBSAs (core-based statistical areas). The building blocks of these defined urban areas are counties. The classifications have been criticized for being confusing and irrelevant to issues of population and urbanism; nonetheless, the county is the finest spatial resolution at which myriad kinds of data are regularly and systematically collected. This is a cogent reason the Census Bureau continues to define many urban areas by using counties as the **minimal mapping unit**. Nonetheless, counties do vary dramatically in their areal extent; for example, San Bernardino County, California, is much larger than the whole state of Rhode Island. Also, many studies demand a finer spatial resolution than counties provide. In fact, many businesses obtain, archive, and analyze data at much finer spatial resolutions than the county level, for example, at the ZIP-code level.

The Census Bureau has tried to keep up with the times by developing finer-resolution definitions for *urban* that are not based on county-level data. This has been possible largely because of the power of computerized GISs. For the 2000 census, the Census Bureau created a more flexible definition of urbanized areas, defining **urban areas** (**UAs**) and **urban clusters** (**UCs**). According to the Census Bureau, an urban area consists of a "densely settled core of census block groups and census

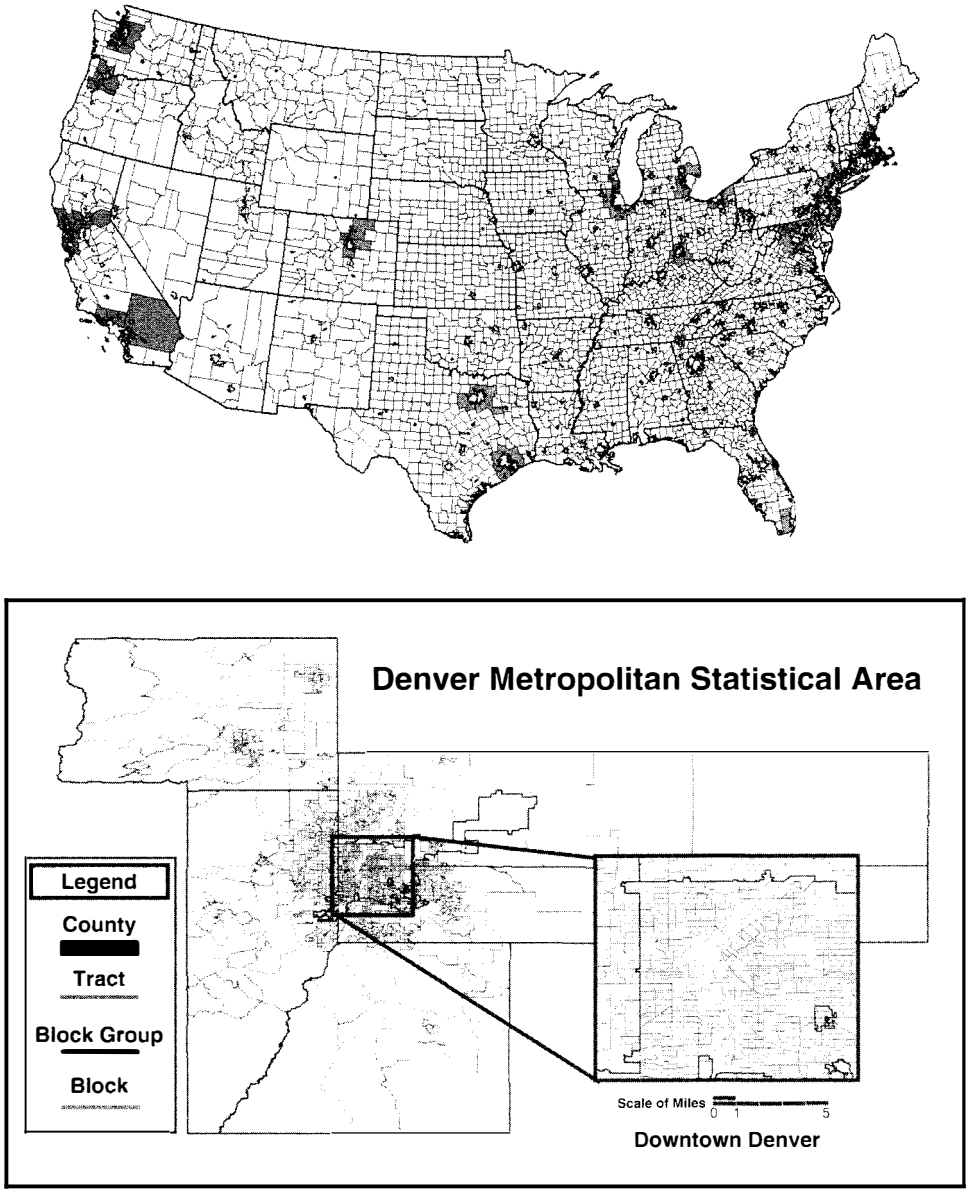

Figure 6.2 Maps showing U.S. census geography for the conterminous United States: (a) States and counties, with urban areas (UAs) and primary metropolitan statistical areas (PMSAs) shaded. (b) Counties, tracts, and block groups for most of the Denver metropolitan statistical area. Inset shows block-level detail for downtown Denver.

blocks that meet minimum population density requirements, along with adjacent densely settled surrounding census blocks that together encompass a population of at least 50,000 people, at least 35,000 of whom live in an area that is not part of a military installation." An urban cluster is defined in the same way but is smaller; it requires a minimum of 2,500 and a maximum of 50,000 residents. Given such efforts on the part of the Census Bureau, we believe geographers can now accurately

characterize human population patterns much more precisely than ever before. Although it is difficult to link definitions of urban and rural to many types of data that are gathered only at the county level, the finer spatial resolution manifest in the newest definitions of urbanized areas does improve the validity and detail of many studies, such as those on urban sprawl, commutersheds, and exurbia. In fact, all the data gathered at the block and block group can be linked to these new definitions of urban. We appreciate that the Census Bureau maintained their county-level characterizations of urban/metropolitan areas at the same time they created a more precise definition of urban from the finest-resolution data available to them.

Limitations of Explicit Reports

We conclude this chapter by stepping back a little, taking a more critical look at the use and meaning of explicit reports in geographic research. First, we return to the observation that explicit reports are quite popular in human geography. They are so popular probably because they strike geographic researchers as so direct—if you want to know where people vacation, and why, just ask them. If you want to understand the meaning of place to people, just ask them. As we have described, these measures are called explicit for the very reason that people *know* they are responding to a request for information by a researcher, and they respond only with what they think of explicitly. As we noted in Chapter 3, this explicitness both increases and limits the value of explicit reports as types of data. We have considered the value of explicitness; we now consider some of its limitations.

To believe that all explicit report data are completely true and accurate would be naive. There are a variety of psychological, social, and cultural processes that influence memory for personally held beliefs and feelings, and their communication to others. We have already referred to some of the more straightforward of these processes, such as that respondents may give socially desirable answers. Or respondents may do the opposite—attempt to thwart or damage the efforts of the researcher. A malicious, impish, or plain antisocial respondent (yes, we have met them) may lie or attempt to deceive the researcher. Finally, we have already discussed the possibility that interviewers can produce distorted responses, because of either their personal characteristics or the way they administer or interpret report instruments.

The limits of memory are clearly important. Both its accuracy and completeness should be questioned. Of course people forget. But even more, psychological research has shown that memory is an elaborative process, a *construction* of what happened as much as a reconstruction or simple retrieval. The emotionality of events can influence this, and not always in the way many people believe (namely, that intensely emotional events are recalled more completely and realistically). Both the recency of the time period asked about and the amount of aggregation of information required by a memory question influence the accuracy of responses. For example, a common type of report item asks people to say how often they do something. The accuracy of responses varies as a function of the time frame over which people are asked to aggregate their activities. Keeping the time frame short and recent leads to more accurate reporting. "How many times do you take the bus each week?" is better than "How many times do you take the bus each month?"

There are some subtler and, we think, more interesting limits to the utility of report data than just the fallibility of memory. The cognitive science of language is a core concern here—how language encodes and expresses one's mental contents; we touched on this in Chapter 5, and we return to it in Chapter 12, where we discuss implications of the "ontology" of geographic features for geographic information systems. This is a complicated issue with a long history of scholarly and scientific debate that has by no means been resolved in the various disciplines that address it (such as anthropology, linguistics, psychology, and artificial intelligence). But we can make a few fairly noncontroversial observations about language, meaning, and mind. The first is that mind does not equal language, and not all meaning and experience can be expressed verbally. Language does express meaning, of course, but in a surprisingly complex and unobvious way; luckily, we don't need to understand how language works in order to use it effectively. Words label categories of knowledge or belief, but despite lots of attention by smart people, we still do not know exactly how. For example, language expresses meaning in a profoundly contextual way—the context being provided by the current situation, past situations, cultural experience, and more. Translation across languages is therefore impressively difficult, and it still has not been completely automated. It is no surprise that altering the wording of items in explicit-report instruments can produce different responses, even when they seem formally equivalent. "Do you think the federal government should forbid corporations . . ." will not mean exactly the opposite of "Do you think the federal government should allow corporations . . ." At the least, this suggests that researchers should interpret responses to explicit-report items in a relative way; for example, comparisons of responses by subgroups exposed to the same wording can be interpreted, but absolute levels of responses to a particular wording perhaps should not be.

Furthermore, as we have noted, explicit reports require respondents to access their mental states (beliefs, emotions) and externalize them for the researcher. But in fact much mental processing goes on subconsciously or unconsciously; Freud may have overstated our interest in sex and death (then again . . .), but he was apparently right about the great portion of mind that is outside of conscious awareness.[8] A person can't provide an explicit answer to something about which he or she is unaware, and that's apparently quite a lot. For example, we can access what we are thinking about but often not the process by which we arrived at that thought. We can usually answer "what" questions much more faithfully than "why" or "how" questions. For example, ask somebody why he or she thinks natural scenery is so beautiful or how he or she managed to steer the car around a turn without crashing. You may get sputtering or some vacuous restatement of the question like "because it makes me feel at one with nature" or "by turning the wheel." We humans are funny creatures insofar as ignorance and chaos make us very uncomfortable.

[8]Theories and debates about conscious versus unconscious mental processing are a big part of the history of cognitive and behavioral science in the 20th century. Besides Freudian psychodynamic ideas, there is introspectionism, behaviorism, and various versions of cognitivism. Modern cognitive psychologists recognize distinctions between automatic and controlled processes, and "implicit" and "explicit" memory.

The illusion that we understand something is very comforting. So when asked to explain something, we often offer rationalizations or personal theories, all the while deceiving ourselves and the researcher that we are directly accessing the mental activities that actually explain a belief or behavior.

It should also be remembered (hopefully explicitly!) that most geographers who study attitudes, beliefs, and memories do so because of what it can tell them about behaviors—what people do and will do, and why. But there is by now extensive literature from social psychology, sociology, and political science that explicit mental states and behaviors have a surprisingly tenuous correspondence. We cannot go into this here in great detail, but the common-sense belief that we do what we do entirely because of what we believe just isn't very true. A sampling of factors that actually help explain behavior besides beliefs in the domain in question include cultural norms and expectations, habits, action opportunities and constraints, personal gain or loss, past experience, and beliefs that one's actions are public or anonymous.

Let's consider a specific example of all this. As we noted above, contingent valuation surveys have been used in an attempt to put a monetary figure on how much people would be "willing to pay" for various aspects of "nature." These surveys can be (and have been) severely criticized, however. Such a survey of 1,000 Americans might show that they would be willing to pay $28 for the continued existence of the grizzly bear, $33 for the bald eagle, $19 for the bison, and so on for other "charismatic mega-fauna." Summing the willingness to pay for enough species would eventually exceed the average household income of the respondents. Another criticism of contingent valuation surveys is their failure to account for people's lack of understanding of what they are valuing. For example, researchers are often interested in the value of "services" provided by various types of ecosystems or biomes. For example, swamps and bogs, now affectionately known as "wetlands," provide many valuable services to humans such as flood control, fish hatcheries, wastewater absorption and cleansing, and so on. Ordinary citizens rarely appreciate the true cost of paving over a wetland for a parking lot. Nonetheless, these costs often come back to haunt citizens in the form of capital losses from disastrous floods, regional economic problems associated with depleted fisheries, and capital expenditures like sewage treatment plants. In other words, a survey of ordinary citizens regarding the value of a local wetland may be a very inaccurate measure of its true value.

Our discussion of the limitations of explicit reports like contingent valuation is not meant to dissuade you from using them in research. Explicit reports are much too useful for that, even with their limitations. We do hope you will think carefully about using them and consider other approaches to data collection. Contingent valuations, for example, can be complemented with measures of what people actually spend on trips to national parks or on homes that have unspoiled views of nature, a technique known as **hedonic valuation**. We favor using explicit reports along with other types of data; any research idea is rightly more believable when it has multiple types of data supporting it, not just multiple data sets. And when you use explicit reports, exercise caution and restraint in interpreting them. For example, responses may be instructive *relative* to responses to other items or in other situations, but they may be of doubtful value in the absolute. Appreciate that explicit reports are more likely to tell you something accurate about some things than about others.

Review Questions

- What are the following types of explicit reports: surveys, interviews, sociometric ratings, activity diaries, contingent valuation, focus groups, protocol analysis, tests?

Format of Explicit Reports

- How do closed-ended and open-ended items differ, and what are some strengths and weaknesses of each?
- How do standardized and nonstandardized items differ, and what are some strengths and weaknesses of each?
- What are some specific types of closed-ended items? What are some specific formats for rating scales?
- What are arguments for and against treating rating-scale responses as metric data?

The Administration of Explicit Reports

- What are options for administering explicit reports, and what issues need to be considered when deciding how to administer them?
- What are strengths and weaknesses of using the Internet to collect explicit report data?

Designing and Generating Explicit Instruments

- How are items for explicit reports created and revised?
- What are some desirable characteristics of explicit-report items?
- How do order effects arise, and what can be done to minimize them?

The U.S. Census

- What is a population census, and how has a census historically been administered and used in the United States?
- What are strengths and weaknesses of the U.S. census as a source of data for scientific research?
- What are the various spatial units in which U.S. census data are organized?

Limitations of Explicit Reports

- How are the quality and usefulness of explicit-report data compromised by the following issues: memory, truthfulness, open disclosure, the linguistic encoding of meaning, access to cognitive processes, and the mind-behavior relationship?

Key Terms

activity diary: type of survey in which respondents record their regular activities; also known as an activity log

adjective or **activity checklist:** type of closed-ended explicit-report item in which respondents mark each alternative that applies from a list of adjectives or activities

American Community Survey (ACS): survey newly created by the Census Bureau in 1996 to be administered yearly to about 1% of the U.S. population; essentially replaces the long form

branching format: follow-up questions that are prestructured to occur whenever a particular response is given to a previous item

census: a count of human population size and characteristics carried out by national governments within a fairly short time period; important secondary source of data for geographers

census block group: geographical unit of the U.S. census consisting of clusters of blocks within census tracts

census block: smallest geographical unit of the U.S. census, consisting of areas typically bounded on all sides by streets, railroad tracks, streams, lakes, and so on

Census Bureau: bureaucracy charged with designing, administering, and distributing the U.S. census

census tract: geographical unit of the U.S. census consisting of small, relatively permanent subdivisions of a county, typically containing about 5,000 people

closed-ended item: item on explicit reports that provides a finite number of specific response options for respondents to choose as answers

context effect: potential influence on responses to particular items on an explicit report that arise because of the context created by items administered previously; a relative order effect

contingent valuation: type of survey in which respondents rate or rank how much they subjectively value something

counterbalancing: a way to deal with order effects by administering all possible orders to equal numbers of respondents; not feasible when certain item orders are required or with large numbers of items

coverage error: generic term for both overcount and undercount on the census

de facto census: an approach to conducting a census in which people are counted according to where they happen to be when they are counted

de jure census: an approach to conducting a census in which people are counted according to where their permanent legal residence is when they are counted

DIME file: Dual Independent Map Encoding file, the standard way to organize U.S. census data within geographic databases since 1980; recently supplanted by TIGER files

explicitness: property of measures in human geography that require people to be aware they are responding in order to provide data to a researcher, and the responses depend on beliefs of which the person is aware

focus group: type of unstructured interview in which respondents in a small group participate in discussions of particular topics

follow-up question: common example of nonstandardized explicit-report item whose specific wording depends on responses to earlier items

forced-choice item: type of closed-ended explicit-report item in which respondents pick one alternative from a list of choices, such as "yes-no," "true-false," or "multiple choice"

free format: follow-up questions that are not prestructured but are formed more or less spontaneously on the spot by an interviewer

generic rating scale: general type of rating scale not designed for a specific purpose or in a specific format such as semantic differentials, Likert scales, and so on

group dynamics: social influences that unfold over time within focus groups and affect what people say and how they say it

hedonic valuation: a measure of how much people value things based on how much they spend to acquire it, used as a complement to contingent valuation

instrument: generic term for a specific example of an explicit report

interview: survey administered and responded to in oral form

items: generic term for questions, tasks, or other units on explicit reports

Likert scale: rating scale on which respondents express their degree of agreement or disagreement with a series of statements

long form: the much larger set of questions concerning more detailed demographic, economic, mobility, and housing information on the U.S. census, administered to a sample of about one-sixth of the population

minimal mapping unit: the smallest spatial unit used to define, measure, and display an areal concept; the U.S. Census Bureau uses counties as the minimal unit for defining cities as "statistical areas"

nonresponse bias: problem of inference in interpreting census data created by the nonrandom subset of the population that does not respond to the Census Bureau's request for information; type of nonparticipation bias, discussed in Chapter 8

nonstandardized items: administering explicit-report items in a way that is neither completely predetermined nor consistent to all respondents

open-ended item: item on explicit report that does not provide a set of specific response options for respondents to choose as answers, but allows them to provide idiosyncratic responses

order effect: potential influence on responses to particular items on an explicit report that arise because of the order, either absolute or relative, in which the item was administered

overcount: threat to the validity of a census arising from people being counted who should not be counted, or people being counted more than once

paired (or triadic) comparison: type of closed-ended explicit-report item in which respondents directly compare entities, two or three at a time, either ordinally or metrically

paired comparison rating: rating scale on which respondents compare entities with the use of a metric rating scale; a metric paired comparison

pilot study: preliminary study designed to practice and evaluate any component of your approach to primary data collection of any type

protocol analysis: type of unstructured and open-ended interview in which respondents "think aloud" about their reasoning during some problem or issue

ranking of alternatives: type of closed-ended explicit-report item in which respondents put two or more alternatives in order from most to least, or least to most

rating scale: type of closed-ended explicit-report item in which respondents provide a number or mark a line to indicate the amount or extent of something, including the degree of belief they have in something

response set: pattern of responses to explicit items that suggest a person was not sincere in his or her responses

sampling debate: controversy over whether the U.S. census should continue to attempt a complete enumeration of the population *or* use sampling and statistical estimation to achieve a more accurate count, potentially

semantic differential: rating scale on which respondents rate a set of entities on each of several adjectives, as a means of expressing attributes of connotative meaning

short form: the small set of questions concerning basic demographic and housing information on the U.S. census, administered to the entire population

social desirability: type of response set that occurs when people provide responses according to what they think is appropriate or socially positive

sociometric rating: type of survey in which members of a group express beliefs about each of the other members of the group

sorting task: type of closed-ended explicit-report item in which respondents sort phrases, pictures, or other stimuli into groups

standard metropolitan statistical area (SMSA): way the U.S. Census Board first defined cities in 1950 on the basis of county-level data; now supplemented or superseded by a variety of related concepts, including MSAs, consolidated MSAs, primary MSAs, and core-based SAs

standardized items: administering explicit-report items in a predetermined and consistent format to all respondents

survey: explicit report in which respondents answer questions about their opinions, attitudes, preferences, activities, or demographics, typically in written form; also known as a questionnaire

test: explicit report in which respondents answer questions about their knowledge, and the researcher is interested in the correctness of the answer

TIGER file: Topologically Integrated Geographic Encoding and Referencing file, currently the standard way to organize U.S. census data within geographic databases; recently supplanted DIME files

undercount: threat to the validity of a census arising from people not being counted who should be counted

unidimensional: desirable property of explicit report items that ask about only one idea rather than two or more

urban area (UA): new way the U.S. Census Board defines cities with data at a finer resolution than county level, instead defining them on the basis of densely settled block groups and blocks that encompass at least 50,000 people

urban cluster (UC): new way the U.S. Census Board defines cities with data at a finer resolution than county level; same as urban areas but smaller, requiring a minimum population of only 2,500 people and a maximum of 50,000

Bibliography

Barrett, R. E. (1994). *Using the 1990 U.S. Census for research.* Thousand Oaks, CA: Sage.

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55,* 803–832.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge: MIT Press.

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist, 59,* 105–117.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.) (2000). *The science of self-report: Implications for research and practice.* Mahwah, NJ: Lawrence Erlbaum Associates.

U.S. Census Board Web site: *http://www.census.gov*

Weeks, J. R. (2005). *Population: An introduction to concepts and issues* (9th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Weisberg, H. F., Krosnick, J. A., & Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis* (3rd ed.). Thousand Oaks, CA: Sage.