CHAPTER 5

Behavioral Observations and Archives

Learning Objectives:

- How are behavioral observations and archives used to collect data in geography?
- How are behavioral observations and archives examples of nonreactive measurements, and how are they not?
- How is scientific behavioral observation different than everyday behavioral observation?
- What is "coding" open-ended records, and how are coding systems developed?
- How are validity and reliability established for coded data?

This chapter covers the two types of data collection known as behavioral observation and archives, used primarily by human geographers. Theories in human geography often involve ideas about human activity in space and place—what people do, what they have done, or what they will do. This includes the behavior of both individuals and various groups, such as families, communities, corporations, bureaucracies, and so on. A straightforward way to collect data on people's current or ongoing activities is to observe and record those activities, which is the technique of behavioral observation. If we want data about people's past activities, we could use archives. Of course, we cannot directly collect data about what people will do in the future, but we can at least use some kind of interview or survey to ask them about their intentions. Chapter 6 covers such explicit reports. We group behavioral observation and archives together in this chapter primarily out of convenience. Besides the fact that both types of data collection are used mostly in human geography (with some applications in physical geography), they frequently share another important property, along with the physical trace measures discussed in Chapter 4: They can both produce data without requiring people to intentionally and knowingly supply information to a researcher. That is, they are often good examples of the nonreactive measures we introduced in Chapter 4, at least when they are collected in such a way that people are unaware they are producing the raw material for data.

Behavioral Observations

Behavior is overt, potentially perceptible action or activity by people or other animals. It is *not* thoughts, internal feelings, purposes, or motivations. That is, behavior is what we do, not why we do it. Behavior is nearly always goal-directed, however, and some behavioral scientists don't consider aimless movement to be behavior. Behavioral observation is a type of data collection in which scientists watch and listen to behaviors of individuals or groups. The behaviors are coded into data during observation, or more commonly, **records** of the behavior are somehow made for later coding into data. Of course, we all observe human behavior all the time, including our own. However, scientific observation of behavior is a bit different. It aims to apply the typical values of scientific practice, such as systematicity and objectivity, to observing behavior. For example, scientific observation should be based on a planned strategy. At a first pass, it should produce data on behavior rather than inferences about the meaning or intention of the behavior. It should include a complete record of the setting and its inhabitants. It should involve an attempt to overcome subjectivity in observation, such as when an observer is more or less sensitive to the occurrence of particular actions because of his or her personal interests.

We noted above that behavioral observation often produces data nonreactively. The degree to which scientific behavioral observation is actually nonreactive varies quite a bit, of course. That depends mostly on whether people are aware they are being observed and recorded. Observers or recording devices can be hidden so the data collection is completely nonreactive. Or the observer or device may be in plain view, but time is allowed to pass before behavior is actually recorded so that those being observed become adapted to the observation and revert to their "normal" behaviors. The most extreme form of this may be **participant observation** in a particular setting, wherein the researcher essentially joins an ongoing setting or subcultural group by becoming a member (yes, this raises ethical questions that we return to in Chapter 14). Finally, we should repeat what we noted in footnote 5 of Chapter 4, that reactance—when people change their responses because they know they are being measured or observed—does not *necessarily* happen even when people are aware they are being observed (we wonder if performers or celebrities only become themselves when they are being observed).

There are a variety of ways to observe behavior. You might simply write **anecdotal records** about what you observe, although this typically lacks the systematicity we look for in scientific observation. Alternatively, you could fill out checklists or rating scales of activities you observe (see Chapter 6). A detailed and comprehensive record of what an individual or group does over some time period is called a **specimen record** (or "running record").[1] Such a record is not usually very useful to a researcher; it is typically best to focus on particular behaviors or particular events. That is, you will probably want to observe less than all possible behaviors at all possible times. This calls for a form of sampling.

You can sample observations whenever you have the time and inclination, but that approach is also lacking in the systematicity that scientists favor. There are two main approaches to **formal observation schedules** that systematically sample from some population of possible behaviors that could be measured. The first is **time sampling.** To do time sampling, choose specified uniform time intervals occurring at specified times, such as one hour every day at 5:00 p.m. Observe and record the behaviors of the people you are studying during those intervals. We discuss sampling designs in detail in Chapter 8, but for now, consider that either randomly or systematically chosen time intervals could be best. Time sampling is appropriate for frequently occurring behaviors, but clearly issues of temporal scale in choosing the size and frequency of intervals are important. The second formal approach to behavioral sampling is **event sampling**, appropriate for situations or behaviors that are rare and would not be expected to necessarily occur in any given time interval. To do event sampling, choose an event to observe, such as a hurricane. Observe and record the behaviors of the people you are studying whenever this event occurs during the duration of the study (one could also sample events during the study duration). Issues of concern when doing event sampling include defining what the relevant event is and establishing that the observation period of the study is adequately representative of some larger population of events. In fact, the concepts of time and event sampling are really quite broadly relevant to many types of geographic data collection used to study processes that occur over time. All geographers must think about sampling from time intervals or events whenever they use data that depend on temporally recurring phenomena.

Behavioral observations can be made and coded live and *in situ,* with an observer or set of observers who are present at the place and time of the behaviors of interest, and trained to code what they see and hear as data. Usually this is not the method of choice, however. It is difficult or impossible to always be there when you want to observe, and to be able to code reliably on the spot in real time. Instead, some type of media is used to make records of behavior, usually to be coded later in order to produce interpretable data (we discuss coding below). There are a variety of media that can be used to record behaviors. Each has strong and weak points, typically with respect to the nature of the record they create and their cost and efficiency. Possibilities include still photographs, audio recordings, video recordings, or some combination thereof. Base maps on sheets of paper or computer tablets can

---

[1]A classic example of a specimen record comes from Barker, R. G., & Wright, H. F. (1951). *One boy's day.* New York: Harper. The authors and their assistants closely followed a child from the time he woke up in the morning until he went to bed in the evening, taking detailed notes on what he did, where he went, who he was with, and so on. Yes, he brushed his teeth before he went to school.

be used to record the locations of activities over time. There are various automatic recording technologies, such as radio or GPS tracking (we discuss this satellite technology in Chapter 12 and the ethical questions about tracking people as part of research in Chapter 14). Transportation researchers use automatic **traffic recorders** that are pneumatic tubes or inductive loop detectors laid across roads and highways. Depending on their specific design, they can automatically generate a count of vehicles or even record their speed of travel. Devices placed in tiles on the floor can record where people step in public buildings. Of course, odometers in vehicles or connected to bicycles can certainly record distances traveled. Another example that is becoming increasingly popular is to automatically record what pages or files people view, and for how long, on computer systems such as the Internet. These **transaction logs** may be considered examples of archives (discussed below) if they are recorded previous to the study, for other purposes.

Behavioral observation is an important tool for data collection by geographers, but it has its problems. As we mentioned above, observation records must usually be coded to become interpretable data. Below, we discuss how labor intensive and difficult it so often is to code well. Behavioral observations also have the problem that the presence of observers (or their equipment) can affect the behavior of those observed. This is not limited just to reactance. Participant observers, to take a case in point, typically become part of a social situation and have an influence on it like anyone would, even if the people being observed have no idea the participant observer is a researcher recording data. Observation also suffers from the subjective and selective nature of perception. It is hard to train observers or coders to overcome their strong tendency to interpret the world meaningfully, to see what they expect or what makes sense to their prior worldviews. And both observers and recording devices literally have *points of view*—they always observe or record from some locations and positions, and not others. As a fan of sports knows, even with extremely expensive recording efforts (think of American football's Super Bowl), some behaviors just don't get recorded or observed. It really doesn't matter how many cameras and how much slow- or fast-motion technology you have at your disposal, you will still miss something. It seems that the ideal observational tool and procedure does not exist, and probably never will: a clear, continuous, and precise record of all the behaviors of all the actors, taken as if you were appropriately near or far simultaneously from all perspectives, without any trace of you being there at all!

Before we turn to archives as data, we note that besides observing humans, researchers observe the behavior of nonhuman animals.[2] Behavioral observation is

---

[2]Reactance and other forms of interactional artifacts (Chapter 11) can be relevant to those who study animal behavior, just as they are to those who study human behavior. An example we enjoy is the famous case of "Clever Hans," the numerically adept horse. During the early 20th century, Hans and his owner traveled around Europe, demonstrating Hans's amazing ability to tap out answers to arithmetic problems and the like posed by his owner. In fact, careful observations over several months revealed that Hans was responding to the subtle body language of his interrogator. The interrogator would lean forward slightly as Hans began tapping and then straighten up when he reached the correct final tap. We still think that's a pretty darn clever horse, even if he couldn't add.

a well-developed approach to data collection by researchers who study animal behavior, such as zoologists and psychologists. Some biogeographers do this too, but it is unusual given the geographer's typical interest in the numbers and locations of animals more than in their micro-behaviors.

Archives

In Chapter 3, we introduced archives as a type of data collected by geographers. Archives are existing records that were not collected for the purpose of a particular geographer's research. More often than not, they were not collected for any research purpose at all, but for various storage, journalistic, or record-keeping purposes. In some cases, archival records were created and kept for idiosyncratic and personal reasons—why does a person save letters and photographs? As such, archives are, by definition, clear examples of secondary data (introduced in Chapter 3). Geographers use financial records, birth and death records, marriage records, law enforcement data, hospital and clinic data, newspaper stories (in any medium), industry and business records, museum records, historical documents, diaries, letters, government records, movies, literature, advertisements (again, in any medium), voting results, tax and cadastral records, retail and wholesale data, and more.

Archives tend to provide a very nonreactive source of data. As secondary sources, they do not reflect any reactance to the measurement activities of a particular researcher. They are sometimes based on procedures in which people were asked to supply information intentionally to someone other than a researcher. Thus, archives can potentially suffer from types of biases that occur when people attempt to create particular impressions with the data they provide instead of responding truthfully. We can't think of a better example than using tax records as data; although we assume our readers are honest citizens like the two of us, a researcher would obviously be gullible to assume that all responses on tax forms were strictly truthful and sincere.

We use the term "archives" to refer specifically to records expressed in symbolic form—typically words or numbers but also images. That is, archives are records that indirectly refer to entities or processes, rather than records that directly express entities or processes. Physical materials (discussed in Chapter 4) that have been stored for some time are not archives in this sense, even if they are sources of secondary data insofar as they were not stored for the purpose of a particular study or researcher. Physical materials such as rock samples or stuffed birds are often stored in museum collections, for example. Of course, researchers use such collections as physical materials for their studies; they come to the museum and make physical measurements on the entities in the collection. We prefer to think of using collections in this way as physical measurement on secondary materials rather than as archives.

Let's consider an example of the use of archives. Early in the history of geographic thought, people's written (and drawn) records of their experiences traveling to foreign lands provided an informal source of information about the peoples and places they had visited. More recently, travel journals and diaries have been more systematically mined as an archival source of data to study the motivations and experiences of long-distance international migrants. And diaries or journals can provide information about many other phenomena too. Comments about rainfall in the journals or logbooks of explorers and expedition leaders have been analyzed as a proxy measure of historical weather patterns, for example.[3] This example shows that physical geographers as well as human geographers can use archives, even relatively unstructured and qualitative archives like diaries. Physical geographers and other earth scientists have in fact used a variety of archives as sources of data, such as ancient texts or inscriptions, governmental and private records, shipping records, and business and commercial records.

Like behavioral observations and responses to open-ended explicit reports (Chapter 6), archives must often be coded in order to produce usable data. Archives are records of activities, events, or characteristics that often cannot be directly interpreted by scientists; they must be processed first. For example, an explorer's journals probably do not give daily rainfall reports in inches or centimeters. This information has to be determined by classifying what was written in the daily entries in terms of categories of more or less precipitation, or none at all. That is, the researcher has to code the journals to extract interpretable data from them. We turn to the subject of coding next.

Coding Open-Ended Records

Both behavioral observation and archives often produce relatively open-ended records of phenomena that must be processed further in order to generate systematic data for analysis and interpretation. A video of pedestrians is a record of their behavior, but it is not in itself data. Likewise, travel journals kept by migrants are not data until their contents are "digested" for meaning. Other types of data collection similarly produce these open-ended records, notably some of the approaches to explicit reports we consider in Chapter 6. An audio recording of a park ranger talking about the way human visitors interact with the flora and fauna of the park is not scientific evidence until it can be systematically interpreted. This interpretation is essentially a process of categorizing the records or their parts. It is more than just the process of physical measurement discussed in Chapter 4. Unlike most of the physical measurements, open-ended records typically consist of words, pictures, or intentional acts that have meaning—they are semiotic or symbolic artifacts created by an

[3]Michaelsen and Larsen discuss the use of archival sources to estimate historical precipitation. They analyzed the journals of John C. Fremont and others in order to reconstruct Southern California precipitation during the 1840s. Michaelsen, J. C., & Larson, D. O. (1991). Historical documentary evidence for Southern California precipitation variability. In M. R. Rose & P. E. Wigand (Eds.), *Proceedings of the Southern California Climate Symposium: Trends and extremes of the past 2000 years* (pp. 93–117). Technical Report No. 11. Los Angeles: Natural History Museum of Los Angeles County.

entity with agency (as discussed in Chapter 3). As such, these records must be *interpreted* in order to extract their basic meaning. Not all human behaviors are semiotic in this way, of course. But even observations of the most ingenuous and straightforward acts usually require extra work to turn them into generalizable scientific data, just as the physical materials of Chapter 4 must be measured to produce data.

The process of turning open-ended records into data is called **coding**. When the records are semiotic artifacts such as verbal or graphical expressions, the process is often called **content analysis**. It consists of two parts, segmentation and classification. **Segmentation** is breaking the records into appropriate units. What counts as "appropriate" is primarily a matter of conceptual and theoretical consideration. What are you interested in? What type of entities is your theory concerned with? Are they words, phrases, sentences, paragraphs, or entire documents? Or perhaps they are images, objects, or parts thereof. In the case of behavioral observation, you might need to know about specific or average weekly acts (such as trips). These considerations are essentially about the analysis scale of your units (Chapter 2). There are some important implications of whether you use narrow or broad units. As in all areas of geographic research, it is optimal if your analysis scale matches the scale of the phenomenon you are studying. As a rule, it is more difficult to generalize with narrow units. What do I know about your travel behavior when I observe that you took one left-hand turn? Broad units may be too coarse or insensitive. What have I missed about your travel behavior if all I know is the total number of miles you travel each month? It is important to recognize that narrow units can usually be aggregated into broader units, but not vice versa.

Once you have determined the segmentation appropriate for your research and explicitly described the process by which you will create the segments, you need to decide how to classify your segments. **Classification** is grouping segments into abstracted categories—assigning segments to the categories—that capture aspects of the content or meaning of the records, aspects that are relevant to the theoretical and conceptual purpose of your research. Classification is virtually always an act of reduction insofar as a certain number of segments are translated into a smaller number of class members, usually many fewer. In rare circumstances, classification could lead to an expansion of data, if each segment is classed into several different categories.

We emphasize that developing a coding system is usually very time-consuming and difficult work. If you find otherwise, it might be because you have very circumscribed records that can be transformed rather straightforwardly into data, which essentially means your records are not very "open-ended" to begin with or you are working with a set of records that contains little variability. We believe that, in many cases, quick and easy coding is a sign that you have developed a poor coding system that will produce data of doubtful value. No matter how easy you find the process of developing your coding system, however, you must make sure to document the final system you use in detail in a **coding manual**. That will prove invaluable in training coders, in describing your procedures when you tell other people about your research, and in allowing people to evaluate or replicate your work.

Categories for coding come in many diverse forms. Their most important characteristic is that they effectively and efficiently capture aspects of the records relevant to your theoretical interests. Generally, categories should be specifically and operationally defined; they should deal precisely with your type of records and your research issues, and they should make it as easy as possible for coders to do their job well. Categories must be **exhaustive** by providing a coding option for every possible segment a coder will encounter in the records. More often than not, categories should also be **mutually exclusive**, with each segment going into exactly one category, although there are definitely situations where it makes senses to allow multiple categories for each segment or even multiple category systems for each segment (our example coding problem below demonstrates both of these situations). It's great if you can devise a coding system where every segment fits into a meaningful category, but usually it helps to include a "miscellaneous" or "other" category for eccentric segments; have coders write brief explanations for each such segment they code as "other." The nature of your records and research question obviously determines the proper number of categories to include in your system. As a general rule, however, you will usually find that three to 10 categories do the job well—more than that may be unwieldy to use or confusing to interpret.

What about the meaning of your categories? That should depend on prior conceptual and theoretical knowledge you and the rest of the research community in your problem domain have accumulated. It should depend on what specific research question or hypothesis you want to address. If you are studying homeowners' beliefs about the likelihood of a wildfire in their neighborhood, you'd better make sure to create one or more categories that concern statements about wildfires happening. Perhaps in a fantasy world of prior omniscience, this conceptual/theoretical or **top-down approach** would be sufficient to dictate your category design. We don't live in that world, however, so one nearly always combines this top-down approach with an after-the-fact **bottom-up approach** to developing categories in which you examine records to get ideas for what categories are needed. In fact, it is not uncommon for categories to be defined almost exclusively by the bottom-up approach of developing categories from the records themselves, as part of the ongoing process of actually coding. In the end, all records must be uniformly coded according to the final category system, of course. Perhaps most often, the top-down and bottom-up approaches are combined in an iterative process. That's part of the reason that developing good coding systems is such hard work.

Example Coding

A detailed example will help illustrate the coding of open-ended records. Our example is based on open-ended explicit reports (Chapter 6) generated as part of some research on how people look at and remember information from topographic contour maps and natural landscapes.[4] The studies examined how perception and memory vary as a function of the amount of training people have had in the use of

[4] The study was conducted by the first author and his colleagues: Montello, D. R., Sullivan, C. N., & Pick, H. L. (1994). Recall memory for topographic maps and natural terrain: Effects of experience and task performance. *Cartographica, 31*, 18–36.

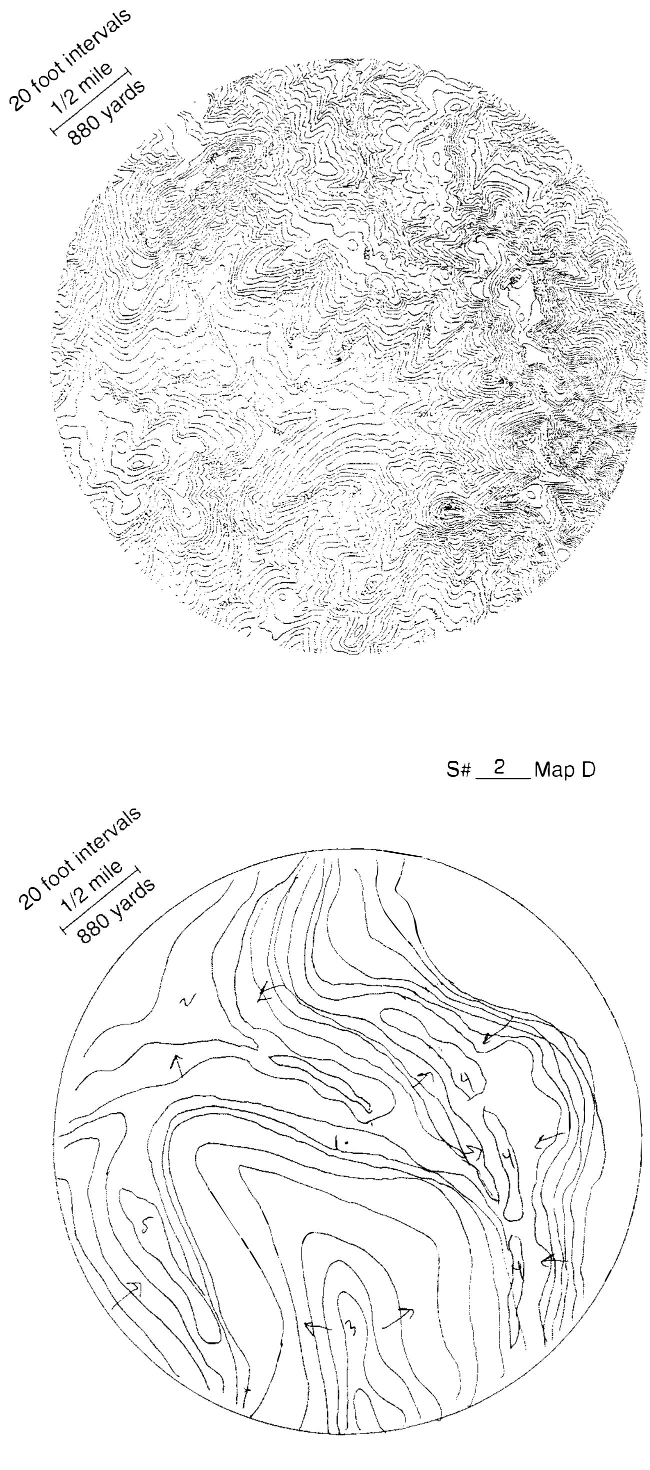

Figure 5.1 Example of open-ended coding based on a study of memory for topographic maps[4]; (a) topographic map portion D, (b) sketch map drawn from memory of Map D by an example research subject.

topographic maps, and as a function of whether the people are required to perform a task matching the map with the corresponding landscape. In one of the studies, research subjects viewed and recalled portions of topographic maps; half of the subjects subsequently matched a heading direction to one of three photographs of landscape scenes they would see if they were standing at the center location represented on the map. Subjects recalled the map portions by verbally describing them into an audio recorder. Subjects also sketched the map while describing it, explicitly labeling features with numbers, in order to help disambiguate the parts of the map being described.

Figure 5.1 shows the topographic map portion referred to as "Map D" (it happens to be from central New Mexico near Albuquerque). It also shows the sketch map of D drawn by one subject (labeled subject 2). Here is a transcription of what this subject recalled about Map D:

> Ok, #1, I'm standing on top of a hill that's right above me. And it's connected to a hill that's on the top left on the circle. Which means pretty much it's the predominant mountain of the map. Umm, below me there's a ravine that comes up from the bottom. Ok, that's #2. #3 is the ravine that comes up and goes to the left. But it's still below where I'm standing. Umm, to the top then right, well more to the right of me #3, no #4, are a bunch of little mountains or little peaks that form like a half circle surrounding the ravine and me where I'm standing, and their slopes are pretty steep to the right. #5, there's a little mountain to the right. I mean to the left and to the bottom, and it's not as high or as steep as #2 or #4. (Experimenter: Anything else you remember?) That's about it.

Box 5.1 shows the coding sheet used to classify elements of the verbal recalls. The researchers were interested in the nature of the features subjects would use to discretize the landscape, and how subjects would describe spatial properties of those features and spatial relations among them. The first step in coding this record was to segment the text into discrete features. This was relatively easy because subjects were requested to explicitly label each feature with a number. In a few instances, however, subjects did not employ the labeling system correctly, and thus features identified by coders did not match the subject's numerical labels. In our example, five features were described:

1–hill (single maximum, also self-location)

2–hill (single maximum)

3–ravine (single minimum)

4–bunch of mountains (multielement maximum)

5–mountain (single maximum)

Then spatial terms were coded. These were considerably more difficult to code than the features, because discriminations among general classes of terms could be quite subtle, and because redundant coding needed to be avoided (redundancy was very common in these transcripts). In our example, 21 spatial terms were mentioned nonredundantly. They are tallied and explained in Box 5.1.

Coding Reliability and Validity

Now that you have turned your records into *coded* records, you have data. As scientists, you want to know whether your coded values really measure what you think or claim they do. After all, other scientists will certainly want to know this when they read your work. Given the complex and sometimes vague nature of your records and your coding system, can you blame a person for being skeptical that your data, your evidence, really reflect the constructs you claim to be measuring? We're not saying that you intentionally miscoded or misrepresented your data, although that would definitely be a problem. We're talking about a system of coding that is too difficult, too underspecified, or too vague to produce adequate measurement. That is, you will want to establish, as much as possible, that your data-coding system is "reliable" and "valid." And you will want to be able to convince others of this.

Chapter 11 is all about reliability and validity. For now, we'll simply define reliability as the *consistency* of measurement; validity, specifically "construct" validity, is the *truthfulness* of measurement. The construct validity of coded records is largely determined by interpreting the apparent meaning of a detailed and transparently documented coding system, and by the statistical relationships of coded data to other types of data that are already understood to be valid. You can establish coding reliability by having two or more coders redundantly and independently code subsets of your records. These independent codes of the same records will largely agree if the coding is reliable. In Chapter 11, we label this "inter-rater reliability." Agreement among coders is usually greater when dealing with specific, concrete, and narrow categories. You and I can agree pretty reliably that a magazine advertisement included a tree in it; we would agree less reliably that the ad was trying to convince consumers that driving a certain brand of automobile would stimulate their "sense of adventure." But high reliability is pretty important for scientific data, as we discuss more in Chapter 11. So another reason that developing a coding system is hard work is that you need to put in considerable effort to refine and improve your coding system and its documentation so that you can achieve satisfactory coding reliability. You can do this by performing intermediate reliability checks on your coding system as you develop it. Have two coders code a small subset of your records with your system as it is, then compare and discuss disagreements. Make changes in the coding system to avoid the disagreements. Repeat until you get reliability high enough; we consider what that is in Chapter 11. The topographic map recalls in our example above were eventually coded with about 80% agreement on feature types and 70% on spatial terms.

To the degree that you want to take a scientific approach to your research, it is a good idea to establish coding reliability and include that in your research write-up. Particularly if your research is meant to be confirmatory rather than exploratory,

Box 5.1 Coding Sheet and Coded Terms for Example of Open-Ended Coding

Topographic Map Memory Study[4]

Coder SAS   Date of Coding Nov 11, '91   Subject# 2   Map D

1. Number of Features 5
2. Number of Feature Mistakes 0
   A feature is wrong, not present, or located incorrectly relative to other features (ordinally).
3. Terrain Feature Types (nouns or noun phrases):
   Code each feature by letter according to how it is first mentioned. Do not code redundantly.

	Single	Multielement
up (maximum)	1, 2, 5	4
down (minimum)	3	
flat		
slope		
mixed/combined		
water		
other		
self-location (feature identified as subject's location; may be redundant)	1	
indeterminate		

4. Spatial Terms (words or phrases):
   Count nonredundant spatial terms not directly prompted by experimenter. *Metric* is quantitative, whether precise or approximate. *Topological* is nonquantitative, expressing order, connection, adjacency, containment, etc. *Ego-dependent* is a spatial term that depends for its meaning on the person's location (on map or in lab). Cardinal directions are typically ego-independent unless person is explicitly mentioned (e.g., "west of me") and topological unless qualified by term like "precisely" or "directly." Do not code "here," "there," "then," "next," or "in this area."

	Ego-independent	Ego-dependent
feature-locational		
metric distance		/
metric direction		
topological/nonmetric	//	///// ///// /

feature-intrinsic properties			
size or number	///	elevation	
slope		metric	
metric	/	topological	/
topological	/	shape	/

Spatial terms marked with gray highlighting:

Ok, #1, I'm standing on top of[a] a hill that's right above me[b]. And it's connected to[c] a hill that's on the top left[d] on the circle. Which means pretty much it's the predominant[e] mountain of the map. Umm, below me[f] there's a ravine that comes up from the bottom[g]. Ok, that's #2. #3 is the ravine that comes up and goes to the left[h]. But it's still below where I'm standing. Umm, to the top[i] then right[j], well more to the right of me #3, no #4, are a bunch of little[k] mountains or little peaks that form like a half circle[l] surrounding[m] the ravine and me[n] where I'm standing, and their slopes are pretty steep[o] to the right[p]. #5, there's a little[q] mountain to the right. I mean to the left[r] and to the bottom[s], and it's not as high[t] or as steep[u] as #2 or #4. (Experimenter: Anything else you remember?) That's about it.

a-topological, ego-dependent
b-metric distance, ego-dependent
c-topological, ego-independent
d-topological, ego-dependent
e-size or number
f-topological, ego-dependent
g-topological, ego-dependent
h-topological, ego-dependent
i-topological, ego-dependent
j-topological, ego-dependent
k-size or number
l-shape
m-topological, ego-independent
n-topological, ego-dependent
o-metric slope
p-topological, ego-dependent
q-size or number
r-topological, ego-dependent
s-topological, ego-dependent
t-topological elevation
u-topological slope

we recommend subjecting a subset of your records to redundant coding by two independent coders. If you can show that inter-rater reliability is satisfactorily high, it will help convince reviewers of your research that you are measuring something with quality and dependability. And once you achieve a suitable level of reliability in your coding system, you need to make sure to rigorously train coders in its use. Also, check your data periodically throughout the coding process to make sure you maintain high coding reliability.

Since our example study was originally conducted in the 1980s, several commercial software packages have become available to facilitate the coding of open-ended records—so-called qualitative data analysis software such as *ATLAS.ti* and *NUD*IST*. These packages do help; they automatically update category definitions and coded units whenever you make a change to the coding system, for example. But don't misinterpret the promise of these packages—they cannot do the hard work of actually interpreting what someone's verbal utterances mean (their semantics). To a large extent, that is, the hardest part of coding cannot be automated. That stems from the fundamental nature of extracting and classifying the semantics underlying words. In most studies, you want to code what verbal text *means*, not what it literally *says*. But the relation of words to meaning is a highly complex problem that is anything but straightforward; it still has not been completely solved by scientists who study language and cognition. The problem is that words relate to meaning according to a "many-to-many mapping" that is typically disambiguated by the context of the situation, of the recent conversation, of the people involved, and so on. A single word means different things, often many different things, and the different meanings can be quite unrelated. In our coding example above, sometimes subjects used the word "below" to mean at a lower elevation in the landscape, and sometimes they used it to mean lower on the page of the map. In many languages, English being a prime example, a single concept can be expressed by several different sets of words, often many different sets of words. The concept of "urban area" may be captured by words like city, town, municipality, central city, village, hamlet, urbanized area, suburb, metropolis, polity, burg, conurbation, metropolitan area, and megalopolis—and that's not even counting technical census terms like "standard metropolitan statistical area" or "urban cluster" (see Chapter 6). These facts about the way all natural languages (as opposed to formal languages such as mathematics) express meaning make coding a difficult task that is hard to perform reliably. They also ensure that there will always be work for scholars of law and of religion, in case any of you were worried about that.

Review Questions

Behavioral Observation

- How does scientific behavioral observation differ from the everyday behavioral observation that we all do?
- What are some different ways to record behavior for future coding and analysis?
- What are the formal observation schedules of time and event sampling?
- What are some strengths and weaknesses of behavioral observation as a technique for collecting data?

Archives

- What are some examples of archival data sources in geography? How do archives differ from physical materials (Chapter 4)?
- To what extent are archives examples of nonreactive measures?
- What are some strengths and weaknesses of archives as a source of data?

Coding Open-Ended Records

- What types of data sources in geography require open-ended coding and why?
- What are the two parts of coding called segmentation and classification?
- What are properties of a good coding system, and how does one go about developing such a system?
- How does one establish the reliability of a coding system?

Key Terms

anecdotal records: nonsystematic form of behavioral observation

behavior: overt, potentially perceptible action or activity by people or other animals

bottom-up approach: generating coding categories by reasoning from an examination of the records that are to be coded

classification: part of the process of coding open-ended records in which segments are assigned to categories that capture aspects of the content or meaning of the records

coding: process of turning open-ended records into data by classifying them, or segmenting them and classifying their parts

coding manual: detailed documentation of your coding system for open-ended records

content analysis: term for coding when the records are semiotic artifacts such as verbal or graphical expressions

event sampling: specific formal observation schedule in which events are recorded at any time they occur during the course of data collection; appropriate for rare events

exhaustive: necessary property of coding systems in which there is a coding option for every possible segment in the records

formal observation schedules: plan for systematically sampling events from a population of events distributed over time

mutually exclusive: property of coding systems in which there is one and only coding option for each segment in the records; usually but not always desirable

participant observation: technique of behavioral observation in which a researcher becomes a member of an ongoing setting or group in order to observe it

records: representations of structures and processes, including human activity, encoded in words, pictures, or sounds; they must be coded or otherwise measured in order to produce scientific data

segmentation: part of the process of coding by breaking records into appropriate units

specimen record: detailed and comprehensive record of what an individual or group does over some time period; also called a running record

time sampling: specific formal observation schedule in which events are recorded during specified time intervals occurring at specified times

top-down approach: generating coding categories by reasoning from prior conceptual or theoretical knowledge

traffic recorders: detectors laid across highways that automatically record the passage of vehicles

transaction logs: records of actions a user takes while interacting with a database or other computer system

Bibliography

Boehm, A. E., & Weinberg, R. A. (1987). *The classroom observer: Developing observation skills in early childhood settings* (2nd ed.). New York: Teachers College Press.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (2[redacted]). *Unobtrusive measures* (rev. ed.). Sage classics series #2. Thousand Oaks, CA: Sage.