1

Introduction to Observational Methods

Observing behavior – the central concern of this book – is an ancient human endeavor without which even our survival could become problematic. What will the beast we hope to stalk, kill, and bring back to the tribe do next? Is that attractive and suitable mate open to my advances? Is that child in trouble and in need of our help?

Not all questions modern researchers pose will be as dramatic as these, and as behavioral scientists search for answers, self-conscious, systematic observational methods will come to supplant raw observation. But what exactly do we mean by observational methods? A definition is in order. In an expansive vein, the eighteenth-century historian William Douglass wrote, "As an historian, every thing is in my province" (1760, p. 230). Similarly, following the nineteenth-century physiologist Claude Bernard (1865/1927), the present-day behavioral scientist could say: Everything I know and do begins with observation. I observe and describe the gait of the horse. I observe and record the infant's weight. I observe whether my participants check *strongly agree*, simply *agree*, or some other choice on a questionnaire.

This chapter is intended as a basic introduction to observational methods. In it we introduce concepts and terms that will be familiar to readers with some experience of observational methods, but that nonetheless provide a foundation for the chapters that follow.

SYSTEMATIC QUANTITATIVE MEASUREMENT VERSUS QUALITATIVE NARRATIVE

Clearly, a definition of observational methods that includes any and all observation colonizes too much territory – although some students arrive on the first day of our observational methods courses thinking that observation only involves looking and then creating narrative descriptions. True, insightful

and informed narratives have a long and important history in such fields as history, journalism, and anthropology, and what are usually called qualitative methods have contributed to a number of fields in the behavioral sciences (see Cooper et al., 2012). Moreover, as we describe in the next chapter, qualitative methods do play a role when developing the coding schemes used for systematic observation. For example, Marjorie Shostak's *Nisa: The Life and Words of a !Kung Woman* (1981) provides an excellent example of qualitative methods at work. In it she organizes interviews around such themes as earliest memories, discovering sex, first birth, and motherhood and loss; and she provides texture, nuance, and insight that would largely elude quantitative approaches. Another classic example is Barker and Wright's (1951) *One Boy's Day: A Specimen Record of Behavior*, which provides intimate and poignant minute-by-minute, morning-to-night observations of one boy's life during a single mid-twentieth-century Kansas day.

In contrast, as we understand the term, observational methods for behavior are unabashedly quantitative. They provide measurement. Measurement is usually understood as the act of assigning numbers or labels to things (Walford, Tucker, & Viswanathan, 2010). In principle, the thing measured could be any discrete behavioral entity. In observational practice, that entity is typically an *event* or a *time interval* within which events can occur (see Chapter 3). As you will see in subsequent chapters, *event* is a key term – we apply it to both relatively instantaneous behaviors and behaviors that have appreciable duration. Some authors – for example, Altmann (1974) – reserve the term for more momentary behaviors and use *state* for behaviors of greater duration.

Measurement implies a measuring instrument: A thermometer gauges a person's temperature, a scale a person's weight. For systematic observation of behavior, the measuring instrument consists of coding schemes – which we discuss at length in the next chapter – used by trained observers. As you will see, unlike more familiar measuring devices, coding schemes are more conceptual. They are based on mental distinctions and not on physical materials like thermometers and rulers, and they involve a human component (i.e., the observers). Melvin Konner's work (e.g., 1976) with Harvard's study of the !Kung in Botswana in the late 1960s and early 1970s provides an example. An electronic device delivered a click to his ear every 15 seconds. He then recorded which of several mother, infant, adult, and child behaviors defined by his coding scheme had occurred since the last click. One result of his work was a quantitative description of how often others in the environment (e.g., mothers, fathers, other adults, siblings, other children) paid attention to and played with !Kung infants.

Measurement also implies a measurement scale. The distinctions we usually make were introduced by S. S. Stevens (1946) some time ago. He categorized measurement scales as: (a) *nominal* or categorical – the names assigned to the entities of interest have no natural order, like agreeable, extroverted, open; (b) *ordinal* – the integers assigned to entities can only be ranked or ordered, like first, second, and third in the race; (c) *interval* – an increment anywhere on the scale involves the same amount of whatever is measured, but zero is arbitrary, like degrees Celsius; and (d) *ratio* – every increment on the scale denotes an identical amount and zero indicates truly none of the quantity measured, like kilograms or many of the summary statistics for individual codes we describe in Chapter 8. As you will see in the next chapter, the coding schemes of observational methods typically rely on categorical measurement.

Perhaps the best way to distinguish the methods described in this book from observation generally would be to call them *systematic*. Thus when we refer to observational methods, it is systematic observation we have in mind. Systematic differs from more informal observation in a number of ways. First and foremost, it involves preplanning. Research questions and key underlying constructs are articulated, and coding schemes developed (see Chapter 2), with the research questions and constructs in mind before observation begins. Observers are then trained, with special attention paid to their accuracy (see Chapters 5 and 6) and the strategies they use to code behavior (see Chapter 3). As Bakeman and Gottman (1997) summarized the matter, central to systematic observation is (a) the use of coding schemes that have been defined and piloted beforehand (b) by trained observers of demonstrated reliability. At heart, it is this approach to measurement that makes observational methods systematic.

CORRELATIONAL VERSUS EXPERIMENTAL DESIGNS

In the world of scientific investigation, measurements are embedded in research designs. A key distinction is between correlational and experimental designs. With correlational designs, values of variables (i.e., constructs) are simply measured (like a person's gender or self-esteem), which allows only weak or no causal inference. In contrast, with "true" experimental designs, values of key variables are manipulated, which allows causal inference. For example, a confederate could be instructed to display either a fear expression or a happy expression during a session, thereby manipulating the type of emotion to which a participant is exposed. In common use (e.g., *The New York Times*), the word *observational* is often used as synonymous with

correlational. Perhaps for this reason, students sometimes think that observational methods are inherently correlational, but this is not so. True, many experimental studies are performed in laboratories and behavioral observations are often employed in field settings not involving manipulation. But correlational studies can also be performed in laboratories and experimental ones in the field; and behavioral observations can be employed for either type of study in either setting. It is the design that makes a study correlational or experimental, not the measurement technique used.

PREDICTOR VERSUS OUTCOME VARIABLES

Whether or not values of some variables are manipulated, another key distinction is between *predictor* and *outcome* variables, which in the context of experimental studies are often called *independent* and *dependent* variables. Other terms are possible; for example, when studies posit more complex causal models, variables whose presumed causes are unspecified and lie outside the model are called *exogenous*, whereas other variables are called *endogenous*.

Typically, but not necessarily, observational methods are used for measuring outcome or endogenous variables for both experimental and correlational studies. As detailed in later chapters, observational variables often detail how much or how often some behavior occurred or whether behaviors were contingent. Often investigators want to know next whether these outcomes were affected by (or associated with, for those who eschew causal language) such predictors as gender, age, diagnostic group, environmental context, type of teacher or instruction; or, in experimental studies, whether they were affected by values of some manipulated variables. Thus in both experimental and correlational contexts, observational methods are often used to determine values for those variables that the investigator hopes can be accounted for by other variables of interest.

VARIABLES, UNITS, AND SESSIONS

Variables attach to something and a useful term for that something is *analytic unit*. As we plan an investigation, describe it for others, and think forward to subsequent data analysis, it is important at the outset to specify two key components: not just our basic analytic units but also our *research factors*. This is true whether or not observational methods are used to determine values for some or all of our variables. Research factors are usually described as *between-subjects* (e.g., gender with two levels, male and

female) or *within-subjects* (e.g., age with repeated observations at 1, 2, and 3 years of age). Between-subject analytic units are, for example, the individual participants, parent-child dyads, families, or other groups (often called *cases* in standard statistical packages, *subjects* in older literature, or simply basic *sampling units*), whose scores are organized by our between-subject research factors. When repeated measures exist, additional *analytic units*, each identified with a level of a repeated measure, are nested within cases.

When observational methods are used, the *observational session* almost always serves as the basic analytic unit, where a session is defined by a sequence of coded behavioral events for which continuity can generally be assumed (although either planned or unplanned breaks might occur during an observational session). Summary statistics and indices derived from the coded data for an observational session constitute scores. Scores from the various sessions (i.e., analytic units) are then organized by any between- and within-subjects factors and are analyzed subsequently using conventional statistical techniques as guided by the design of the study.

Typically an observational study involves two steps. First, either behavioral events or time units within which events may occur are coded for a session. As noted earlier, this usually involves nominal measurement; although as discussed later, rating successive segments of a session or the entire session using ordinal scales is another possibility. Second, summary scores are derived from the coded nominal data for the session. These scores represent variables of interest and attach to the session. Such scores usually represent equal-interval ratio-scale measurement (e.g., the summary frequencies and other statistics described in Chapter 8) and – taking into account whether variables are between- or within-subjects – can be analyzed (assuming appropriate distributions) with standard statistic techniques such as correlation, multiple regression, and analysis of variance.

In sum, systematic observation is simply one of many possible measurement methods. In common with other methods, systematic observation provides scores for subsequent statistical analysis. In fact, it is common for scores in any given research project to derive from a variety of methods – for example, gender and age from a questionnaire, maternal depression from a self-report scale, and maternal responsiveness to her infant's cries from systematic observation. What distinguishes observational from other methods is that, unlike questionnaires in which responses to a manageable series of questions are elicited, observation is carried out by trained observers who typically code behavior over relatively long sessions. As a consequence, behavioral observation is often quite time-consuming. When coding live, observers need to be present during sessions that can vary from

a few minutes to several hours. More typically, sessions are recorded, which can absorb even more time as observers spend hours coding just a few minutes of behavior. Compared to the few items of a typical self-report measure, data collected from observation can be voluminous and their analysis seemingly intractable. Why then would an investigator bother with such a time-consuming method?

WHY USE OBSERVATIONAL METHODS?

There are many good reasons for using observational methods, but we believe three are particularly compelling (Bakeman & Quera, 2012). First, when research participants cannot tell us what they think or when they cannot read and respond to questionnaires or when they cannot make entries in a diary – as is true of preverbal infants, preliterate children, and animals generally – observational methods provide a way to measure indirectly what is "on their mind." Thus it is not surprising that many early classic examples of observational research involved animals and human infants (e.g., Altmann, 1965; Parten, 1932). Moreover, even when our research participants are verbal, observational methods may still be the best choice if the focus of our research is their nonverbal behavior. In fact, in some cases (e.g., marital interaction studies), it may be interesting to gather data by observational methods about how people actually behave, and then compare those data with other data collected by questionnaires or self-reports about how they say they behave.

The second reason is that spontaneous behavior often seems more natural than elicited behavior. *Natural* is a relative and perhaps slippery term, but when research participants whose behavior is not elicited are observed – and it does not matter if it is in laboratory or field settings – we assume that their observed behavior reflects their proclivities and untutored repertoire. We do not make similar assumptions when the behavior is elicited by the experimenter, for example, when asking a participant to fill out a questionnaire. Participants might be asked to soothe a crying infant in a contrived setting, but somehow the behavior we then observe seems more natural than responses made to a questionnaire asking how they would soothe a crying infant. Nonetheless, we may still wonder whether behavior is changed by being observed – like observer effects in physics. The answer seems to be that humans habituate rapidly to being observed. For example, as reported by Bakeman and Helmreich (1975), marine scientists living in a space-station-like habitat fifty feet below the surface of Coral Bay in the Virgin Islands were on-camera continuously; yet as they went about

their work, awareness of the cameras seemingly disappeared within the first several minutes of their two- to three-week stay in the habitat.

The third reason is that when investigators are interested in *process* – how things work and not just outcomes – observational methods have the ability to capture behavior unfolding in time (which is essential to understanding process) in a way that more static measures do not. An important feature of behavior is its functionality: What happens before? What next? Which are causes and which consequences? Are there lagged effects between certain behaviors? Only by studying behavior as a process can investigators address such questions. A good example is Gottman's work on marital interaction (1979), which, based on characterizations of moment-to-moment interaction sequences, predicted whether relationships would dissolve or not. Also, process questions almost always concern *contingency*. For example, when nurses reassure children undergoing a painful procedure, is the children's distress lessened? Or, when children are distressed, do nurses reassure them more? In fact, contingency analyses designed to answer questions like these may be one of the more common and useful applications of observational methods (for details, see Chapters 9 and 11).

SEQUENTIAL ANALYSIS OF BEHAVIOR

The third reason just given for using observational methods – an interest in process – motivates much of this book. Understanding process means looking at behavior in sequence as it unfolds in time, but – although the terms *sequential analysis* and *observational methods* both occur in this book's title – not all studies that are observational are sequential. The difference is perhaps best conveyed by examples. Three paradigmatic studies that illustrate how observational studies may or may not be sequential were cited by Bakeman and Gottman (1997). These studies all involved preschool children observed in relatively natural contexts and are worth revisiting.

The first study is Mildred Parten's (1932) study of social participation among preschool children conducted at the University of Minnesota's Institute of Child Welfare in the late 1920s. During the 1926–27 school year, some forty-two children whose ages ranged from not quite two to almost five years of age were observed seventy different times, on average. The daily observations occurred during indoor free play and lasted 1 minute for each child; the order in which children were observed varied so that the 1-minute samples for each child would be distributed more or less evenly throughout the hour-long free-play period.

Code	Definition
Unoccupied	Child not engaged with anything specific; seems aimless.
Onlooker	Child watches other children playing, but does not join in.
Solitary or independent play	Child plays alone and independently, seemingly unaffected by others.
Parallel activity	Child plays independently beside, but not with, other children but with similar toys; no attempt to control who is in the group.
Associative play	Child plays with other children, with some sharing of play materials and mild attempts to control who is in the group.
Cooperative play	Child plays in a group that is organized for some purpose, for example, playing house or a formal game or to attain a goal.

Figure 1.1. Parten's (1932) coding scheme for social engagement.

Parten was interested in the development of social behavior in young children. Accordingly, she asked observers to code each 1-minute sample by the level of social engagement that predominantly characterized it. Her six codes are detailed in Figure 1.1. From the coded intervals, Parten computed the percentage of samples assigned each code, separately for each child. Over the school year, each child was observed for only 70 minutes, on average. Still, her sampling plan let Parten use these percentage scores as estimates of how much time each child devoted to a particular level of social engagement during free play that year. In turn, this let her evaluate hypotheses such as that older children would spend more time in associative and cooperative play than younger children.

However, her data do not let us ask how any of these play states were sequenced in the *stream of behavior* (to use Roger Barker's [1963] felicitous phrase). We cannot determine, for example, whether *Parallel* often preceded *Associative* and *Associative* often preceded *Cooperative* play, not because Parten's codes are not up to the task but because her recording method – coding daily, isolated 1-minute samples – does not capture sequential information. This is not a criticism of Parten – her research questions did not require examining moment-by-moment sequences of behavior. Instead, our intent is to make the point that when sequential data are collected, not just questions like Parten's, but a whole other array of interesting questions can be addressed.

Parten (1932)	**Smith (1978)**	**Bakeman & Brownlee (1980)**
Unoccupied	Alone	Together
		Unoccupied
Onlooker		
Solitary		Solitary
Parallel	Parallel	Parallel
Associative	Group	Group
Cooperative		

Figure 1.2. The evolution of three similar coding schemes for social participation as discussed in the text (adapted from Bakeman & Gottman, 1997).

The second paradigmatic study is provided by Peter Smith (1978). Parten's study had established an association between age and social participation: As children became older, they tended to participate more at higher levels. As ordered in Figure 1.1, each code suggests a higher level of participation than the one before it, so it is tempting to view her codes as suggesting a developmental progression in which parallel activity is a stage through which children pass as they develop from solitary to social group players; that is, Parten's coding scheme could be viewed as an ordinal scale of social participation and not just a categorical one. Smith, however, sought to test that notion of developmental progression directly. For our present purpose – asking what makes a study sequential – his study is useful not so much for what he found out as for the way his modification of Parten's method challenges our sense of what we mean by a sequential analysis.

Simplifying Smith's (1978) methods some, he reduced Parten's six codes to three (see Figure 1.2). He wanted to test explicitly the idea that parallel play is an intermediate stage in social development. As a result, there was no need to distinguish between the presumed precursor stages of *Unoccupied*, *Onlooker*, and *Solitary*. Consequently, he lumped these three into a single code, *Alone*. Likewise, there was no need to distinguish between *Associate* and *Cooperative*; he lumped these two into a single code, *Group*. Smith's recording method was similar to Parten's: He used a sampling strategy to code brief, isolated intervals for the forty-eight children in his study. From these coded intervals Smith computed for each child the percentage of samples assigned each code, separately for each of his study's six successive five-week periods. Then, the code with the highest percent score became the code assigned to the entire five-week period. Examining these coded

sequences of six five-week periods, Smith reported that many children moved directly from a five-week period in which *Alone* predominated to one in which *Group* play did without an intervening period during which *Parallel* play was most frequent. Note, however, that Smith's results can mask the fact that periods in which only *Alone* and *Parallel* occurred (but *Alone* predominated) could be followed by periods in which only *Parallel* and *Group* did (but *Group* predominated); by dividing time into shorter periods, *Alone*-to-*Parallel*-to-*Group* transitions might have been revealed.

Smith's question was sequential as was his analysis, although at one step removed from most examples we give in this book. He used information derived from nonsequential behavioral coding to then code much longer segments of time (five weeks), whereas most examples we present in this book – and the sense in which we usually use the term *sequential analysis* – code moment-by-moment, event-by-event sequences of behavior.

The third paradigmatic study is Bakeman and Brownlee's (1980) study of parallel play. Parten seemed to suggest that parallel play characterized an obligatory development phase, whereas Smith suggested the phase might be optional. This discussion caused Bakeman and Brownlee to think that the question itself might be misleading and that parallel play might better be regarded not as a stage, but as a strategy – important because of how it was positioned and functioned in the moment-by-moment stream of children's play behavior. Therefore, they posed what is clearly a question of behavioral sequencing.

Like Smith, Bakeman and Brownlee (1980) modified Parten's codes (see Figure 1.2). They kept Parten's and Smith's *Parallel*, Parten's *Solitary*, and Smith's lumped *Group*, but they lumped Parten's *Onlooker* with *Unoccupied* and created a distinct new code (*Together*) defined as essentially unoccupied with a focus on others, but without the focus on objects or activities required for *Parallel* and *Group*. Forty-three three-year-old children were video-recorded for about 100 minutes each during free play over several mornings of a three-week summer camp. Observers then viewed the recordings and coded successive 15-second intervals using the scheme just described.

Later we will have more to say about Bakeman and Brownlee's (1980) method of interval recording and will explain why we regard it as less than optimal, but for now we will assume that their data provided a reasonably accurate estimate of how the play states (the codes representing levels of social participation) defined in Figure 1.2 were sequenced in time for each child. Using techniques explained in Chapter 9, Bakeman and Brownlee counted how often various play states followed each other