

5-24-2010

Behavior and Behavior Assessment

Christopher Agnew
Purdue University, agnew@purdue.edu

Janice Kelly
Purdue University, kellyjr@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/psychpubs>

 Part of the [Psychology Commons](#)

Recommended Citation

Agnew, Christopher and Kelly, Janice, "Behavior and Behavior Assessment" (2010). *Department of Psychological Sciences Faculty Publications*. Paper 31.
<http://docs.lib.purdue.edu/psychpubs/31>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Behavior and Behavior Assessment

Janice R. Kelly & Christopher R. Agnew
Purdue University

Draft: 5-24-10

Chapter prepared for the *Oxford Handbook of Personality and Social Psychology*, M.

Snyder & K. Deaux (Eds.), New York: Oxford University Press.

Abstract

This chapter addresses the questions of 1) what do we mean by “behavior” in personality and social psychology, and 2) how can we best assess social behavior. We define behavior as being observable and socially meaningful, but also discuss the dimensions on which behavior varies (e.g., intentional vs. habitual, discrete vs. continuous). We also discuss important variabilities in behavior as they relate to issues of measurement (e.g., behavioral frequency or desirability). For behavior assessment, we focus on some of the practical issues involved (e.g., choosing a coding system, selecting an observational setting), as well as how behavior assessment might intersect with personality and social psychological theory (e.g., manipulating or assessing behavior as it serves as moderator, mediator, or outcome in a theory). We end by discussing some emerging technologies that might prove useful for behavioral assessment (e.g., virtual reality), as well as a call for more integration of behavioral measures into future research.

Keywords: behavior, behavioral assessment, behavioral observation, research methods

Behavior and Behavior Assessment

The past decades have featured a growing focus on cognition, affect, and motivation within personality and social psychology. With the rise of social cognition has come a reduction in focus on the behavioral consequences of it (Agnew, Carlston, Graziano, & Kelly, 2010). There has been a corresponding waning focus on direct observation of what people do (Baumeister, Vohs, & Funder, 2007). That is, data provided by independent observers who supply systematic descriptions of something they have actually seen someone else do has declined. Funder and colleagues have decried the decrease in behavioral focus among personality psychologists (Furr & Funder, 2007; Furr, Wagerman, and Funder, 2010), as have others, but the trend does not appear to be reversing. This is unfortunate, as behavior is obviously a critical ingredient -- if not *the* critical ingredient -- in understanding people.

Personality and social psychology have a long history of focusing on overt behavior. Some of the most well-known studies in social psychology involve the enactment (or lack of enactment) of key social behaviors. For example, early studies of discrimination by LaPiere (1934) focused on observing the behavior of hotel and restaurant personnel in response to requests for accommodation by a Chinese couple. LaPiere's emphasis was on the discrepancy between expressed negative attitudes toward minority group members versus actions toward them, but his main interest (and one that he extolled psychologists to keep their eye on) was on what people actually did. Understanding the cognitive, affective, and/or motivational underpinnings of discriminatory behavior were certainly (and continue to be) of interest, but this interest was primarily in the service of understanding discriminatory actions themselves.

Without the presence of overtly negative discriminatory behaviors, there would be little motivation to understand the underlying psychology.

Sherif et al.'s (1961) classic Robber's Cave experiments involved the observation of teams of boys who faced team vs. team competitions, as well as staged shared disasters that required inter-team cooperative responding. The behaviors studied were real, with consequences for the participants and with clear parallels to situations beyond Robber's Cave. No questionnaire study could have had the same type of impact on generations of psychology students or on subsequent scholarly thinking regarding intergroup relations.

Darley and Latané's (1968) bystander intervention experiments also centered on actions (or inactions) of people in an emergency situation. Would bystanders help someone in clear need of it? Although the focus was on the factors that increase or decrease likelihood of providing help, helping itself was of paramount concern. Similarly, Milgram's (1975) series of obedience experiments focused on how far participants would go in following the orders of an experimenter. What would people do when told to harm another, under the orders of an authority figure? Likewise, personality psychology has emphasized the actions of individuals from the inception of the field (e.g., Hartshorne and May's, 1928, research on children's character and "honesty" behaviors). However, with the shifting emphasis in social and personality psychology toward the psychological underpinnings of behavior, the focus on behavior itself and on its attendant measurement challenges has left the spotlight.

This situation is particularly ironic given the rise in the general public's interest in and appreciation for the kinds of behaviors social psychologists study. For example,

witness the enormous success of writer Malcolm Gladwell's books in recent years. *The Tipping Point* (2000), *Blink* (2005), and *Outliers* (2008) are replete with examples based on classic and current social psychological research. Moreover, television networks worldwide have a seemingly endless fascination replicating or reenacting a number of classic social psychological studies featuring consequential behaviors (e.g., Milgram's obedience studies and Darley and Latané's bystander intervention studies). If there was ever a time when behavior should be at the forefront of social and personality psychology, it is now.

Of course, there are understandable barriers to maintaining a central focus on behavior. It certainly costs more to observe actions *in situ* than it does to collect self-reports of behavior. Beyond cost, there are also time demands that tend to favor methodologies that do not involve real-time behavior assessment. Moreover, there are many behaviors that are very difficult if not impossible to directly observe (e.g., illicit or low-frequency behaviors). However, as we hope to make clear, the importance of keeping social psychological and personality research grounded in actual behavior cannot be understated.

In this chapter, we consider the issues of what behavior is, what behavior is meaningful to examine from the perspective of social and personality psychologists, and how best to assess that behavior.

What is Behavior?

What is behavior? More specifically with respect to the focus of this volume, what might social and personality psychologists consider to be behavior? Even a cursory review of the literature provides a rich and varied answer to the question. For some, any

action by a person constitutes a behavior of interest. At times, specific behaviors are of primary interest for either basic (e.g., the study of mimicry) or applied (e.g., the use of condoms to prevent HIV transmission) reasons. Moreover, some behaviors are more meaningful than others, either from the perspective of an actor or from the perspective of a researcher.

We take the position that behavior is what a person does overtly. Behavior is, thus, observable. We would not include as behavior unobservable internal processes, such as neural activity or the physiological machinations of brain components (Aron, 2010). We further assume that behavior must be meaningful, on some level, to be worthy of investigation in and of itself by social and personality psychologists. So, for example, although social psychologists who study social cognition may track the keyboard pressing of study participants, they are not ultimately interested in the overt behavior of keyboard pressing. Such behaviors are considered to be critical proxies for cognitive, affective, or motivational processes that are themselves of primary interest. *Thus, we define behavior as overt or observable actions that are socially meaningful.*

Obviously this is a very broad definition that encompasses many types of actions. Meaningful social actions may include nodding, talking, smoking, donating to charity, assembling an AM radio kit, smiling, walking on a treadmill, engaging in mutual eye gaze, or holding your hand in freezing water. However, even within this general definition, important distinctions can be drawn regarding various dimensions of behavior (see Table 1), including its intentionality, whether it is consciously guided, and whether it is relatively independent or interdependent. Such distinctions have implications for measurement.

Insert Table 1 about here

Intentional behaviors, or behavior that follows from one's desire, have been the subject of investigation by social and personality psychologists for decades. Models of decision-making have as their core the assumption that most behaviors are enacted in line with the intentions of the decision-maker (e.g., Theory of Planned Behavior: Ajzen, 1991). In contrast, models of habitual behavior do not assume that intentions guide the enactment of habits (Verplanken, 2010). Although habits may have their origin as intentional behaviors (e.g., carefully buckling one's seat belt when first learning to drive), such behaviors may become increasingly unintentional over time (e.g., buckling up "automatically" upon getting in a car).

Behaviors also differ in the degree to which they may be considered guided by conscious awareness. Bargh and others have documented behavior that is not the result of conscious guidance (Bargh & Morsella, 2010). For example, work by Chartrand and Bargh on behavioral mimicry has shown that people copy (or "ape") the actions of others without conscious awareness of doing so (Chartrand & Bargh, 1999). The adequacy of self-reports to capture underlying reasons for such behavior is particularly suspect (Nisbett & Wilson, 1977). Indeed, how can one be expected to know what motivated an action that one had no awareness of enacting?

Some behaviors are conducted without the co-action of others (e.g., reading, eating alone). Such independent behaviors stand in contrast to the multiplicity of behaviors that are interdependent, requiring the co-action of another person for their

enactment (Agnew, 1999). Interdependent behaviors simply could not be enacted independently (e.g., sexual intercourse). From an assessment perspective, such interdependence has both positive and negative implications. On the one hand, there are multiple reports available concerning interdependent behavioral enactment. On the other hand, multiple actors may not provide researchers with converging data regarding what did or did not occur. The lack of convergence, of course, may in itself be of theoretical interest, but the assessment challenge remains (e.g., Bolger, Stadler, Paprocki, & DeLongis, 2010).

Behavior may also be said to differ based on the number of actions that compose the “behavior.” Winking at a colleague is a discrete (and, perhaps, discreet) behavior. Driving an automobile, in contrast, is a series of interrelated behaviors that combine to form the behavior “driving.” Depending on the research question, any number of the discrete actions that form the overall behavior “driving” may be of principle interest. Specificity with respect to the precise behavior of interest is, therefore, of critical importance from a measurement perspective (Ajzen & Fishbein, 1977). Assessing when a given behavior begins and ends (e.g., work on “break points” by Newtonson & Engquist, 1976) can be a salient research consideration.

Behavior also differs in the degree to which it is discernable from other behaviors or whether it must be embedded in a context in order to be understood. The same collegial wink mentioned above might be interpreted as having vastly different meanings by the “winker” (who, for example, may have meant it to signify to a colleague the connection with matter discussed previously), the “winked at” (who may have completely missed a connection with the earlier discussion and is bewildered by the

wink), and an outside observer (who wonders what is going on between the two). In this case, the behavior itself is discrete and easily described: closing one's eye quickly. But the meaning of the behavior is determined entirely by context and not all participants in the setting featuring the wink are necessarily "in" on the context.

Dimensions Relevant to Measuring Behavior

Behavior is not unidimensional, and with each variation of behavior comes issues important to appropriate assessment of that behavior. Below we outline some of the major variations of types of behavior and discuss some of the associated measurement challenges. These distinctions are summarized in Table 2.

insert Table 2 about here

Novel vs. habitual behaviors.

Initiating a behavior for the first time is different from initiating a behavior that one has enacted in the past. Different factors are likely considered before engaging in a behavior for the first time than are considered once the behavior is well-practiced. For example, donating blood for the first time may give rise to a host of thoughts ("Will this really help someone?") and emotions ("Will it hurt?") that no longer cross the mind of a veteran donor. Moreover, repeated behaviors become habitual and decreasingly subject to cognitive input. Given its fleeting and transitory nature, truly novel behavior can be exceptionally challenging to capture either in the lab or the field.

Low frequency behaviors.

Some behaviors occur so infrequently that certain types of measurement procedures would be inappropriate. For example, for low frequency behaviors, direct observation of the behavior would be impractical given that it is unlikely that a low frequency behavior would actually occur during the timeframe of an observation period. However, some low frequency behaviors are extremely important to understanding human social life. For example, acts of heroism are important in furthering understanding of prosocial behavior. In order to gain knowledge about rare behavior, researchers often must rely on self-report measures of some type, certainly actor report, but perhaps also well informed other reporter's of a person's behavior as well. Biases of self-report, however, might be particularly threatening when the low frequency behavior is also socially undesirable (e.g., criminal behavior).

Socially undesirable behaviors.

Behaviors vary in the extent to which people wish to be associated with their enactment. Although researchers may be interested in learning about the cognitive or motivational underpinnings of littering, finding people who are willing to admit to, for example, littering can be challenging given the negative associations attending the behavior (e.g., inconsiderate of others, bad steward of the environment). Traditional approaches to detecting impression management tendencies include administration of social desirability inventories (such as Paulhus's, 1991, Balanced Inventory of Desirable Responding), the scores on which may be used as a covariate in statistical analyses. However, self-reporting behavior via an inventory is clearly a step removed from evidence of actually engaging in socially (un)desirable behaviors. Of course,

researchers may be willing to “take what they can get,” but refining approaches to capturing less desirable (yet frequently occurring) behavior remains important.

Assessing Behavior

The complexity of behavior is clear. It is therefore understandable that the assessment of behavior may be complex as well. There are four primary methods (physiological measures notwithstanding) for assessing outcome variables in social and personality psychology – self-report, informant report, direct observation, and trace measures (Webb et al, 1981). Some of these methods are particularly suited for assessing particular types of outcomes. For example, informants may be particularly good at providing behavior assessments within a specific context (e.g., teachers assessing behaviors in the school context) or sometimes across various contexts (e.g., a parent or romantic partner who interacts with the participant across a wide range of contexts). And self-reports are currently the only method that can be used for tapping internal cognitive and emotional states. Each of these methods has certain strengths and advantages, but each has weaknesses and disadvantages. For example, self-reports are relatively inexpensive to administer and can be used to assess a wide range of traits, attitudes, beliefs, and behaviors. But they are subject to various reporting biases, such as errors in retrospective memory, socially desirable responding, or response sets (Paulhus & Holden, 2010). Trace measures may be more immune to deliberate biases and distortions, especially given that the traces were not originally left for research purposes, but the number of trace measures that are meaningful for

psychological research are somewhat few (see, however, our later discussion of Gosling et al., 2002).

All four methods potentially can be used to assess behavior. If you are interested in a person's behavior, or if you are interested in forming inferences about a person based on knowledge of their behavior, then you can ask the person themselves about their behavior (self-report), ask an informed other about the person's behavior (informant report), directly observed the person's behavior yourself (behavioral observation), or examine the evidence that is left behind from the behavior (trace measures). With respect to behavior assessment specifically, each method also has strengths and weaknesses. Self-reports of behavior are still subject to various memory and reporting biases. We may not be consciously aware of some of our behavior, such as normative or habitual behaviors, so that we cannot report their occurrence with accuracy. Other behaviors that we think may be associated with pejorative attitudes may be more deliberately underreported (e.g., use of swear words). Third party informants may be able to provide estimates of behavior with somewhat more accuracy than self-reports, but their estimates may be based on knowledge of behavior within a limited context (Spain, Eaton, & Funder, 2000) and/or they may also be subject to some of the same biases affecting the actor (Sande, Goethals, & Radloff, 1988). So, for example, when a child is assessed for ADHD, their behavior must be reported within different and specific contexts, such as parent's report of behavior at home and teacher's report of behavior at school, in order to obtain a more accurate and global picture of the overall behavior (Achenbach, 1991). And some behaviors may be more private and less likely to be observable by a third party. Trace measures are somewhat

specifically meant to unobtrusively measure behavior. For example, wear and tear of tiles may be used to assess foot traffic in a building (Webb et al., 1981). Cialdini's impressive body of work on norms also takes advantage of trace measures of behavior. For example, Cialdini counted the number of flyers that were discarded in various settings as an index of adherence to descriptive norms for disposing of trash (Cialdini, Reno, & Kallgren, 1990). However, many of the behaviors of interest to social and personality psychologists (e.g., conversations, emotional displays, interpersonal interaction) do not leave such observable traces. Direct observation of behavior, on the other hand, can be used to assess most types of important social behaviors, and this observation can be done over a range of contexts or within specific contexts. Because the behaviors are recorded as they happen, those reports are less likely to be distorted because of memory degradation or socially desirable responding.

In this chapter, we have therefore opted to focus on direct observation as the primary method of behavior assessment. We do so for a number of reasons. As described above, direct observation of behavior has a number of important strengths, particularly in terms of eliminating biases in respondent self-reports such as social desirability or errors of memory. But also, we want this chapter to serve as a call for researchers in the behavioral sciences to return to the assessment of actual behavior rather than the measurement proxies that have become so commonplace (Agnew, Carlton, Graziano, & Kelly, 2010; Agnew & Kelly, 2010; Baumeister, Vohs, & Funder, 2007). As Furr and Funder (2007) argue, behavior itself is the ultimate source of knowledge for all psychological information. We have long known, at least since the classic studies by Nisbett and Wilson (1977), that people are often unaware of the

factors that drive behavior. Therefore, without direct observation of behavior, the accuracy of such self-reports is in question. Without the ability to directly access another's thoughts, emotions, and motivations, behavior provides an important indication of another person's psychological state unfiltered by a person's possibly biased self-report regarding the behavior.

Behavioral Observation

We argue that direct observation of behavior is, in many cases, the most valid method we currently have for assessing behavior. If we are interested, for example, in the effects of some intervention on smoking behavior, we can assess those effects by directly observing smoking. If we are interested in how negotiators resolve a conflict, we can simply observe and code their negotiation behavior. Given that we are a social science and that social behaviors are for the most part directly observable, direct observation of behavior would seem to be the method of choice for assessing such social behaviors.

However, direct observation of behavior is less commonly used today than self-reports of behavior, and for obvious reasons. Behavioral observation can be difficult. Behavioral observation generally involves either selecting or designing a system for coding behavioral observations. That is, the behavior in question must be defined, and a system must be developed for systematically and reliably recording that behavior. The system must then be used in real time (whether live or from recordings) to assess the ongoing behaviors of choice and usually involves extensive training of coders in order to reliably implement the system. As opposed to self-reports of behavior, this sort of behavioral observation is generally time-intensive and resource-consuming and those

factors are undoubtedly primary reasons for the lack of direct behavioral observation in social and personality psychology. Behavioral observation can also simply be not practical for assessing some behaviors, especially when we are talking about behaviors that are low frequency, private, undesirable, or illegal.

There is some indication, however, that the use of direct behavioral observation is increasing. There have been a number of calls to arms concerning the field's overreliance on self-report and underreliance on behavioral observation (Baumeister, Vohs, & Funder, 2007; Furr & Funder, 2007; Moreland et al., 2010). These calls stress the importance of studying actual behavior rather than behavioral substitutes (Furr & Funder, 2007). In addition, emerging technologies may be making direct observation of behavior more manageable (e.g., Mehl, 2009).

Furr and Funder (2007) argue that despite the difficulties of direct behavioral observation, behavioral data are fundamental to social and personality psychology. As stated earlier, they argue that behavior is essentially the end state or defining characteristic of all of scientific psychology. Funder (2006) argues that personality psychology involves the understanding of persons, situations, and behaviors and the links between them. However, lack of attention to the latter two elements has led to extensive knowledge about traits, but relatively little attention to the classification of behaviors and the situations in which they occur (also see Holmes & Cavallo, 2010). In fact, both social and personality psychology have suffered crises concerning the lack of empirical evidence linking important, proximal psychological constructs (e.g., attitudes and traits) to the behavioral outcomes they are meant to predict. We are well aware of the outcome of Wicker's (1969) classic review of the attitude-behavior relationship as

well as Mischel's (1968) critique of the relationship between personality traits and behavior. Given those classic examples, as well as more recent critiques (e.g., Agnew et al., 2010), behavior must re-emerge to serve a more central role in comprehensive frameworks of social and personality psychology.

The following sections focus on the mechanics of behavioral observation (see Figure 1 for an overview), given that lack of knowledge of the steps involved in behavior assessment could be a major impediment to utilizing this key research tool. What is important to understand is that the issues involved in the definition of behavior discussed at the beginning of this chapter are essential to keep in mind at each step of the assessment process. For example, as will be discussed more fully below, if we are interested in examining behaviors that are infrequent or novel, we may be more likely to choose to study these behaviors in a laboratory setting with a task that is designed to elicit those behaviors.

Formulating the Research Question and Defining the Behavior

While defining the research question might seem like an obvious step in the process of collecting behavioral data, it is an essential one. The research question that is generated has implications for all further steps in the assessment process. It does not matter whether behavioral measures are used in research if they do not address and inform the researcher's primary research question.

We argue, as do others (e.g., Aron, 2010; Weingart, 1997), that behavior can be included in research questions as mediators, as moderators, or as outcomes. That is, behavior can be assessed as an intervening step between inputs and outcomes;

behavior can moderate the relationship between inputs and outcomes; or behavior can be assessed as the final outcome in a sequence of processes.

A number of theories exist in both social and personality psychology that include behavior as an intervening process. For example, in behavioral confirmation theory (Synder, 1984), the pattern of behavior elicited from a target mediates the relationship between the source's expectations concerning the target and the source's confirmation of those expectations. Bem's (1972) self-perception theory of attitude formation suggests that in the absence of knowledge about our attitude toward an object, we examine our behavior in order to infer our attitude. Too often, though, when behavior is assumed to be a mediating process, it is assessed through self-report, or sometimes merely assumed, rather than through behavioral observation.

Behavior can also serve as a moderator of other perceptions or behaviors. For example, in a powerful demonstration of the effects of behavior on participants' psychological and emotion states, Williams and Sommer (1997) had confederates engage in a ball-tossing game where they either included or excluded (ostracized) a naïve participant. Ostracized participants showed greatly depleted fundamental needs and negative emotional states. As another example, Gaddis, Connelly, and Mumford (2004) had leaders display either positive or negative affect after a failure experience. Leaders who displayed negative affect were judged by followers as less effective and responded with poorer quality performance compared to leaders who displayed more positive emotions.

It is perhaps easier to think of behavior as an outcome of other processes, including internal states and motivations, than as mediator or moderator. For example,

an empathetic focus on a victim may lead one to a behavioral outcome that expresses helping (e.g., donating hours of babysitting, contributing money; Batson, 1998).

Oftentimes, however, our research ends prior to the actual behavioral expression of these processes. For example, we may ask people to indicate on a questionnaire how many hours they would be willing to babysit to help out a person in an unfortunate position, but the actual behavioral expression of this intention is not assessed, and as mentioned previously, decades of research have addressed how the realities of a situation may reduce the likelihood that an expressed behavior will result from an intention (Fishbein & Ajzen, 1975).

It may also currently be difficult for social and personality researchers to generate interesting and meaningful research questions that center around behavior as a mediator, a moderator, or even as an outcome given the field's current focus on internal states such as attitudes and emotions. Part of the reason for this may be illustrated in the examples provided above. Using behavior as a moderator often involves the use of confederates, which can be a resource-consuming approach to research. And of course, assessing behavior as an outcome involves the time- and resource-consuming strategy of behavioral assessment. However, given the importance of behavior in validating other forms of assessment, more research questions need to be generated that include the assessment of behavior as a final outcome measure.

Selecting or Developing the Behavioral Observation System

Researchers who are knowledgeable about behavioral observation make the recommendation that, when possible, new researchers should try to borrow from observational systems that have already been created (Bakeman, 2000; Furr & Funder,

2007). Systems exist to code a wide variety of social behaviors, including emotional displays (FACS coding, Ekman et al., 2002), aggressive behavior (Fagot & Hagen, 1985), personality (RBQ; Funder, Furr, and & Colvin, 2000), and group interaction (SYMLOG; Bales & Cohen, 1979). The ability to utilize an existing system, with an already established track record of producing usable and valid data that can be reliably collected, can put a researcher far ahead in the game of behavior assessment.

If a researcher chooses to develop a novel coding system, the researcher needs to keep two primary factors in mind when selecting what sorts of behaviors to include in the system: (1) the level of analysis at which the behavior is to be observed and measured (molecular or molar) and (2) whether the system is intended to generalize across situations or apply specifically to a single situation (Furr & Funder, 2007). The answers to these questions are often dictated by the researcher's theoretical orientation and specific research question.

Level of analysis.

Behaviors can be observed and measured at many different levels of analysis. For example, traditional behaviorists measured behavior at a very molecular level, such as number of bar presses. Some social psychologists who are interested in the expression of emotion also tend to observe and measure behavior at a molecular level. For example, FACS coding (Ekman, Friesen, & Hager, 2002) involves the coding of micro-movements of facial muscles. Observers of nonverbal behavior are often interested in eyebrow flashes, body orientation, and the use of discrete gestures, all of which would be considered molecular level behaviors. For example, Hall and Friedman (1999) had observers record frequency counts of head nods, smiles, and hand

gestures, as well as touch, forward lean, and interruptions in order to understand the interplay of status and gender in a structured interaction. What is important to keep in mind is that these very specific behaviors are consistent with the level of analysis of the researcher's theoretical approach and specific research question. An advantage of measuring behavior at this very molecular level is that the assessments involved require very little inference on the part of the observer about the social meaning of the behavior. Often the judgment is simply whether the behavior has occurred or not, and if the behavior is easily discerned from background noise, such a judgment should be relatively straightforward.

Behavior can also be measured at extremely molar levels. For example, Bernieri, Reznick, and Rosenthal (1988), in a study of interaction synchrony, had observers make global judgments of how smoothly the behavior of mother-infant pairs was coordinated in genuine interactions vs. artificially generated pseudointeractions. Similarly, the Leadership Trait Questionnaires (LTQ; Northouse, 2007) asks both leaders and subordinates to rate the leader on ten behaviorally-based traits, such as being articulate, perceptive, and self-confident. Note that these more molar behaviorally-based judgments involve a good deal more inference on the part of the observer than do specific discrete behaviors such as "gestures." In fact, a trait such as "self-confident" involves the observation of a number of more specific behaviors, such as participation rates or interruptions, from which position on the trait in question is inferred. And the greater the amount of inference involved, the greater might be the potential problems of interrater agreement (Funder et al, 2000). Consequently, it becomes important for the researcher to carefully specify what behaviors he or she

thinks are indicative of the underlying (and non-observed) trait, as a great deal of slippage can occur between the identification of a trait and its operational definition. However, again, the effort that might be involved in achieving sufficient levels of interrater reliability would be worth it if this level of analysis is appropriate for the researcher's conceptual network and specific research question.

A classic example of a more molar system is the Bales' Interaction Process Analysis (IPA) System (Bales, 1950). The IPA was originally developed to systematically categorize small group discussions in a manner that reflected Bales' theoretical ideas concerning group problem-solving. In that earlier work, Bales proposed that a group moves through a particular sequence of task phases as it moves from the beginning to the completion of a task: 1) orientation (gathering information and clarifying the task); 2) evaluation (assessing that information); and 3) control (deciding what to do). Twelve categories of discussion acts were developed to capture movement through those phases. For example, the categories of "gives orientation" and "asks for orientation" reflect activity in the orientation phase. The categories of "gives opinion" and "asks for opinion" reflect activity in the evaluation phase. Note then that the categories are meant to code the process meaning of verbal (and nonverbal) statement, and not their literal content. The implication of this is that the degree of inference needed on the part of the coder is quite high and intercoder reliability can be difficult to achieve without extensive training efforts.

The Riverside Behavioral Q-sort is an excellent example of a mid-level observation system (Funder, Furr, & Colvin, 2000). The RBQ asks observers to make behavioral judgments such as "appears relaxed and comfortable," "seems to enjoy the

interaction,” or “exhibits a high degree of intelligence.” Similarly, the Leadership Behavior Description Questionnaire (LBDQ; Ohio State Leadership Studies, 1964) asks followers to make behavioral judgments of a leader such as “acts as spokesperson for the group,” “makes accurate decisions,” and “keeps the work moving at a rapid pace.” Although a variety of molecular behaviors may go into the assessment of these mid-level categories (e.g., ‘seems to enjoy the interaction’ might be indicated by a participant smiling frequently and actively engaging in discussion with their interaction partner), the level of inference needed in order to make judgments of mid-level categories is much less than for the types of molar judgments described above. A corresponding decrease in the amount of training required in order to achieve an appropriate level of reliability is also a benefit to assessments made at this level.

Generalizability of the system.

The researcher also needs to decide whether the system will be useful across a variety of situations, or whether the system is appropriate only for use in a single type of situation. For example, the RBQ (Funder, Furr, & Colvin, 2000) was specifically developed to be useful across a wide variety of situations involving interpersonal interactions. Observers who use this system are able to sample a variety of potential behaviors that may underlie a particular RBQ rating. For example, “dominates the interaction” may be indicated by extensive talking in one setting and prominent body language in another, making the system flexible across a variety of situations. Indeed, the RBQ has been implemented in a variety of interpersonal situations (Funder, Furr, & Colvin, 2000). The Bales IPA (1950) or SYMLOG systems (Bales & Cohen, 1979) are also flexible systems that can be used across a variety of situations. The flexibility of the

Bales IPA is due to the fact that utterances are coded for their process meaning, rather than coded for literal content.

In contrast, a wide variety of systems have been developed that focus on the coding of behaviors that are specific to a particular situation. For example, Foushee and colleagues (Foushee, 1984; Foushee et al., 1986) developed a system for observing the behavior of a flight crew within the cockpit of an airplane. Although this system was highly appropriate for the situation for which it was developed, and although the observations provided by the system were highly predictive of flight crew performance, the system has little use in other settings because many of the behaviors observed are not relevant to or are not performed in settings other than airplane cockpits (e.g., uses pre-flight checklist).

Kelly (2000), in a discussion of observation systems that have been used for the observation of group processes, made a distinction between process-focused systems and setting-focused systems. Process-focused systems, such as the Bales' IPA, code the process meaning, rather than the literal content, of verbal utterances. Setting-focused systems may utilize either process or content coding, but are distinguished by having an application to only a specific group-task-situation. The avoidance of literal content makes process-focused systems flexible over many types of groups and many types of tasks, except where such group or task features vary too greatly from the theoretical categories proposed by the system. However, it is sometimes difficult to use these systems to test hypotheses that differ from the theoretical underpinnings of the system. In contrast, setting-focused systems are valuable for analyzing interaction in specific groups in specific situations, but cannot be used with any other kind of groups

or even with the same group in any other performance context since categories are often idiosyncratically defined. The usefulness of these systems for theory building is also limited.

Kelly (2000) also describes activity-focused systems that concentrate on the single dimension of vocalization/silence. These systems do not take into account the literal content of a vocalization, but merely record content-free features of the vocalizations such as tone of voice, turn-taking, interruptions, and so forth (e.g., Chapple, 1970; Dabbs and Ruback, 1987; Jaffee and Feldstein, 1970). The obvious benefits of such systems are the ease and reliability with which such observations can be made and the flexibility of the system in terms of applicability to many group-task-setting situations. An obvious drawback is the lack of conceptual clarity of the meaning of sound/silence patterns.

Choosing the Task and Setting

Once the behaviors of interest have been identified, the researcher must decide on the best setting in which to observe those behaviors. Direct observation of behavior can occur in both natural and contrived (e.g., laboratory) settings. The choice of setting must naturally depend on the research question of interest. But it also depends on the particular behaviors of interest as well.

It is not uncommon to equate behavioral observation with qualitative assessment of behavior as it occurs in natural settings, and although the equation is not accurate, the use of natural settings is certainly one choice that researchers might make. There are a wide range of overt, public social behaviors that are available for observation and that can be meaningfully fit into an appropriate theoretical framework. For example, a

broad range of research examining children's levels of aggression has drawn heavily on behavioral observation of children in schools (Fagot & Hagan, 1985; Ostrov & Keating, 2004; Pelligrini, 2001). The various systems that have been implemented make use of both specific, discrete actions (e.g., hitting), as well as verbal utterances indicative of relational aggression (e.g., Crick, 1996).

Naturally occurring behaviors or classes of behaviors can vary in terms of frequency. Higher frequency behaviors have a higher likelihood of being captured in a natural setting. The assessment of lower frequency or novel behaviors may be aided by the creation of a behavior-eliciting task in the lab to ensure that those behaviors occur. For example, in order to examine aggressive behavior, which for most people is a low frequency behavior, participants were required to engage in a task with a confederate that involved provocation on the part of the confederate, with the participant later being given the chance to retaliate with some sort of aggressive behavior (Buss, 1961). In this way, a low frequency behavior can be reliably elicited in an experimental context.

Training Coders and Establishing Reliability

Proper training of observers or coders is essential to the behavioral assessment process. However, this training process can be one of the most time consuming and costly aspects of proper behavior assessment. Some simple instances of behavioral observation, such as a mere frequency count of a discrete behavior like smiling, require little observer training, and in fact some behaviors can be easily assessed with mechanical counters such as a pedometer. Similarly, a popular behavior measure used by intergroup researchers involves assessing seating distance between participants when the participants are allowed to position their own chair (e.g., Goff, Steele, &

Davies, 2008). Other assessments require extensive training. For example, FACS training requires hundreds of hours of training for each coder (Ekman, Friesen, & Hager, 2002). Similarly, a moderately complex observation system developed for coding group interaction was estimated to require approximately 60 hours of training per coder in order for them to achieve appropriate levels of interrater reliability (Weingart, Hyder, & Prietula, 1996).

The degree of training involved depends on a number of factors including the unitization of the activities being observed and the degree of inference involved in selecting an appropriate coding category. For example, in FACS coding, coders are instructed to distinguish and code extremely minute differences in facial muscle contractions, and correspondingly, training is a fairly laborious process. In contrast, coders using the RBQ (Furr, Funder, & Colvin, 2000) are asked to make mid-level judgments (e.g., expresses criticism, exhibits social skill) of a focal interactant in an interaction sequence. Funder and colleagues estimate that RBQ training requires approximately 2 hours per coder.

When developing a novel observation system, training coders and developing the observation system is often a recursive process. Issues of definition of the behavior, including issues of threshold, interdependence, or discernability often lead to a redefinition of how a behavior is operationalized in the observation system. When differences exist for how a particular behavior is characterized and coded, the resolution of those differences can lead to a clearer understanding of the parameters of the behavior in question.

Good behavioral assessment should also involve some sort of indication of the reliability of the system. That is, two or more observers should observe the same event and code the same behavior in the same way. Different indices of reliability exist for different types of data (e.g., categorical vs. interval). A thorough explanation of all these indices along with their statistical applications is beyond the scope of this chapter. However, LeBreton and Senter (2008) and von Eye & Mun (2005) provide comprehensive coverage of these topics.

Most behavioral data is ordinal or categorical in nature. That is, most systems involve categorizing a discrete behavior into one or more unordered categories. To assess reliability for these data, most indices involve some derivation of percent agreement. Simple percent agreement (total number of agreements in categorizing the behavior divided by the total number of observations) is easy to calculate and is easily understood by most readers. However, total percent agreement can overestimate reliability especially when only a few categories are used for the categorization. Cohen's Kappa (1960) is recommended for assessing reliability for categorical behavior systems where errors can be made in the placement of the observed behavior into a specific category. Cohen's Kappa takes into account not only total agreement, but also corrects for chance agreement in situations where such chance error is likely to occur (e.g., few categories).

Cohen's Kappa does, however, tend to provide a lower than expected reliability estimate when there is an unequal frequency of codes across categories. In this case, a weighted Kappa may be the more appropriate index of reliability.

For interval data (e.g., a frequency count of smiling), most researchers use some variation of a correlation coefficient. However, a Pearson correlation assesses only the relative pattern of responses between coders, and does not take into account absolute agreement. Intra-class correlations (McGraw & Wong, 1996), on the other hand, take into account both the relative pattern of responses and absolute agreement as well. Intra-class produces a coefficient that can be interpreted similarly to a correlation coefficient.

Strengths and Weakness of Behavior Observation

We have previously mentioned some of the strengths and weaknesses of behavioral observation, but some aspects of these qualities merit further reflection. One of the primary strengths of behavioral observation relative to self-reports of behavior is that it tends to be less subject to some sources of bias, such as lack of an accurate memory for events, biasing reports so as to appear in a positive and socially desirable light, and various response sets that may occur when reporting on behavior. Memory biases not only include the degradation of a memory trace, but the fact that much of our behavior is simply not noted at a conscious enough level to be stored in memory. For example, behavior that is routinized or habitual is likely not to be consciously noted or recorded in memory. (Just think of the times when we end up in our regular parking space at work with no real memory of the trip that brought us to that location.) And some researchers have suggested that such habitual behavior may make up as much as 40% of our daily activities (Verplanken, 2010). Furthermore, memories can be re-examined and reinterpreted in order to fit more easily into schemas that we hold for ourselves or others (Loftus & Cahill, 2007).

These points all relate to the most important strength of behavioral observation. Given that we have no direct access to a person's internal cognitions and emotions, our best objective indicator of those thoughts and emotions is what a person says and what a person does. Therefore, the systematic recording of those behaviors gives us particularly useful information about the internal psychological workings of individuals.

But behavioral assessment is not without weaknesses as well. As we have noted, observation systems vary quite widely on factors such as the size of the behavioral unit being observed and the degree of inference needed in order to categorize a particular behavior. In terms of implementation of a system, we have already noted that the amount of training necessary to achieve appropriate levels of interrater reliability varies with factors such as size of the behavioral unit and degree of inference. And although we have referred to behavioral assessment as being "objective," observers can also have expectations or biases that color their coding or interpretation of behavior (Rosenthal & Rubin, 1978). Consider, for example, an observer who is recording specific aggressive behavioral acts enacted by young children. Suppose further that this observer believes that boys are more aggressive than girls. The same specific behavior – a pointed finger, perhaps – might be interpreted as an aggressive behavior when exhibited by a boy and not when exhibited by a girl. That is, the observer's expectations or stereotypes may bias how a specific act is interpreted or fit within a behavioral recording system.

As with any research approach, replication is one of the best tools that we have in our arsenal to overcome weaknesses in research methods. Given the potential strengths of behavioral assessment, and given the increasing availability of tools for

collecting these data, research using behavior may reach a critical enough mass for replications to occur. We end this section on behavior assessment by mentioning some of these emerging tools and technologies that might aid us in this endeavor.

New Technologies in Behavior Assessment

Behavior assessments are often made online or live as the focal person is behaving. This is especially true when the behaviors of interest are relatively straightforward and easily distinguished from other behaviors, such as large gestures, laughter, or talking. However, it is also a common procedure to record an ongoing interaction for later coding. Here, the accuracy of behavioral assessment is only as good as the instrument used, and instruments are subject to failure. Videotaped recordings of behavior may be of poor quality, making coding, especially of verbal behavior, quite difficult. Electronic instruments, such as video-recorders, are also subject to instrument decay or deterioration over time.

Despite the potential problems with using mechanical recordings of behavior, recent years have witnessed the introduction of new technologies that may prove to be particularly useful for behavioral assessment. We review several of these technologies here. Note, however, that some of these technologies do not have extensive research applications to date. They are, thus, presented as potential aids to behavioral research.

The Electronically Activated Recorder (EAR).

The Electronically Activated Recorder (EAR; Mehl, 2009) is a developing behavioral sampling technology and technique that is proving useful in sampling behavior as it occurs in a natural context. The EAR is a pocket sized audio-recorder that periodically records small audio samples of the ambient sounds from a person's

environment. The EAR can be unobtrusively carried in the participant's pocket and audio samples can be collected with minimal amounts of reactivity as the person goes through his or her natural daily activities. Behavioral information is then coded from the sampled audio data. EAR has been used to assess a number of common social behaviors, such as talking (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007), swearing (Mehl, et al., 2006), and class attendance (Mehl et al., 2006). Many personality researchers consider experience sampling to be the best proxy for direct behavioral observation (Spain, Eaton, & Funder, 2000). However, EAR allows for the sampling of behavioral data independent of the participants' self-reports, and thus provides a behaviorally-grounded criteria against which to assess other personality measures (Mehl, 2009).

Virtual reality.

Behavioral assessments can and have been applied within the context of immersive virtual environment technologies (IVET). Earlier we mentioned that it may be difficult for social and personality psychologist to conceptualize research where behavior serves as a moderator. IVETs, however, provide a vehicle for delivering behavioral manipulations. For example, IVETs can be used to manipulate subtle nonverbal cues or other behavioral characteristics of avatars in order to better understand their function in social interaction. A variety of behavioral dependent measures can also be conceptualized, including social interaction behaviors such as social distance between a participant and an avatar, vocal utterances, nonverbal cues such as body orientation or gaze directed toward the avatar, and so forth.

For example, Bailenson, Yee, Blascovich, Beall, Lundblad, and and Jin (2008) manipulated physical proximity between a teacher and a student in order to examine its effects on student learning outcomes. They did so to simulate typical classroom environments in which only a subset of students can be at desks in closest proximity to the teacher. They found that student performance was improved when the student sat in a seat that was closer versus farther from the teacher. In another experiment, Bailenson and colleagues (2008) manipulated the presence of a model co-student or distracting co-student to assess the effect of the presence of such co-students on learning outcomes. They found that the behavior of virtual co-learners influenced learning patterns of participants. Of particular interest, they found that eliminating the presence of co-students altogether had the most beneficial effect on learning.

In earlier work, Bailenson, Blascovich, Beall, and Loomis (2003) used IVET to study the interpersonal distance maintained between participants and virtual humans. They found that distance maintained was influenced by direction of avatar approach (front versus back), avatar gaze behavior (mutual vs. nonmutual with the participant), and violations of personal space. Manipulating environmental conditions via IVET provides a particular powerful means for examining both behavioral influences and consequences.

Webcams, surveillance monitors, and the Internet.

Recordings of public behavior are ubiquitous with the advent of webcams and surveillance recording machines. Security cameras record the behavior of passers by, live webcams broadcast the behavior of selected exhibits on the internet. In fact, cell phones can also easily record the behavior of others in practically any location. For the

most part, the behaviors recorded are public behaviors and thus can ethically be used as a medium for behavior assessment. The internet itself can provide a wealth of information on behavior (Wallace, 2001). On YouTube, for example, people post videos involving many sort of social situations or behavior, and in fact, viewers can post comments or feedback about the posted videotape. The internet can itself be used to assess behavior. With webcams, the behavior of people in distributed locations can be observed and coded.

Online participant observation is also an emerging method for behavioral assessment (Utz, 2010). Online participant observation is especially useful for studying sensitive topics (e.g., discrimination), difficult to access groups (e.g., viewers of online pornography, criminals), and interpersonal interaction (e.g., online dating sites).

Advantages of participant observation include a fairly naturalistic record of behavior if the observer's identity is unknown. Beyond possible ethical issues involved in covert participant observation, researchers are limited in their ability to interpret some text due to lack of information concerning motivation and emotions. Generalizability to other groups or even to similar groups that meet offline can also be limited. Utz also provides a scheme for the types of behavioral data that can be collecting including text messages, avatars, group size, etc.

Social networking websites, such as Facebook, are also now being used increasingly by personality and social psychological researchers to study various questions relevant to behavior. Self-reported behavioral information is a hallmark of such sites, but there are also opportunities for researchers to assess actual behaviors enacted by users with respect to interactions with other users or in obtaining information

about other users. For example, Waggoner, Smith, and Collins (2009) used Facebook to assess differences in impression formation for passive versus active perceivers. Passive perceivers (defined as those Facebook users who were provided with preselected information about target individuals and whose searching behavior was, thus, constrained) were contrasted with active perceivers (those users who were free to choose the type and amount of information they received about a target person). Results indicated that active perceivers, whose behavior was not constrained, liked targets less but were less confident regarding their judgments of target personality than were passive perceivers. Manipulating the behavior of social network website users provides a fascinating new approach to the study of an increasing ubiquitous human behavior: web use.

The challenge for using these technologies is in developing appropriate research questions that could be addressed using available recordings. That is, the researcher must limit his or her questions to those where the recordings would provide meaningful data. The potential of the resource, however, is exciting.

Coder computer support systems.

Coding from digital recordings of behavior has been made easier with the advent of various computer observational support systems, such as the Observer Video-Pro (Lucas et al., 2000), the NVivo (<http://www.qsinternational.com>), or the MAXQDA (<http://www.maxqdo.com>). These systems aid the task of observation by offering direct data entry of data into a computer, often in conjunction with an automatic time stamping of the ongoing interaction. The interface eliminates errors due to transcription and the

time stamping increases the researcher's ability to assess coding reliability. Generally, these systems allow you to import observations from analog/digital videotapes or digital media files. Observation streams can be reviewed and edited on the computer. In conjunction with this, software allows you to define a number of behaviors and assigned a key code to each behavior. These keys are pressed to record the occurrence of an event during an observation session.

Observation of Personal Living Spaces (PLSs).

An intriguing new technique for discerning personality traits through observation has been initiated by Gosling and his colleagues (Carney et al., 2008; Gosling et al., 2002). Rather than directly observing behavior, however, Gosling and colleagues observe the remnants or traces that behavior has left behind by observing personal living spaces (PLSs) such as dorm rooms or offices. The Personal Living Space Cue Inventory (PLSCI; Gosling et al., 2005) is an observation tool that allows the user to comprehensively document aspects of individual's personal living spaces (PLSs), such as whether the space is generally tidy or colorful, whether the clocks are set on time, or whether the space was light or stuffy. These descriptors have also been linked to various personality traits. For example, the Big Five personality trait of Openness to Experience has been linked to PLSs that are distinctive and colorful, whereas the traits of Conscientiousness has been linked to PLSs that are tidy and decorated in a conventional manner (Carney et al., 2008; Gosling et al., 2002).

Integrating Behavior into Social and Personality Psychology Theory

We have argued, as have others, that behavioral assessment is essential for researchers who study personality and social behavior because behavior often provides an objective look into the hearts and minds of research participants. Since we do not have direct access to internal states that are the precursors of behavior, such as attitudes and predispositions, other than as they are processed through other mechanisms subject to potential bias (e.g., self-reports), overt behavior provides the most direct access to these sorts of psychological states. While it is true that we may also be interested in cognitions and emotions for their own sake, knowledge of how internal states impact behavioral outcomes is an essential part of understanding the human condition.

However, merely assessing and quantifying behavior is not enough to provide meaningful insight into human psychology. Those behaviors need to be translated into psychologically meaningful constructs – that is, constructs that are theoretically useful for making sense of people’s psychological states. And further, that behavior needs to be integrated into our social and personality psychology theories.

Interpretation of Behavior as Psychologically Meaningful Constructs

A problem with some aspects of behavioral observation is that the psychological meaning of a specific behavior is often unclear. For example, a single cooperative choice in a Prisoner’s Dilemma Game may be reflective of a person’s underlying cooperativeness. However, it may also reflect a more complex strategy for gaining future resources. For example, a tit-for-tat strategy, where one responds on a subsequent move with what one’s partner has responded on for the previous move, is an effective strategy for establishing a pattern of cooperation. But it is also an effective

strategy for reliably gaining resources. The behavior itself is ambiguous with respect to the motive that underlies its' enactment. As another example, striking out at another person, an overt behavior that should be fairly easy to reliably notice and record, could be motivated by fear, and thus reflect a self protective response, or by anger, and thus reflect an exertion of frustration, power or dominance over the other.

Other examples of psychologically ambiguous behaviors include those often produced by activity-focused systems. As described previously, activity-focused systems are those that focus on the presence or absence of an activity such as talking. Although this single dimension of behavior has been used to generate complex descriptions of personality (Chapple, 1970), the precise psychological motivation behind a specific pattern of behavior is difficult to discern. For example, does rapid and uninterruptable speech indicate that the speaker is extraverted, nervous, or dominant? Are long "turns" of speech also possible indicators of those same psychological states?

In some instances, we put the burden of this translation on the observers by asking them to make at least some degree of inference in assessing the behaviors. Coding systems that operate at the more molecular level, for example, often explicitly ask observers to develop trait or character inferences based on their observation of sometimes long, but sometimes very short (thin slices; Ambady & Rosenthal, 1993), sequences of behavior. In general, the ambiguity of a behavior increases as the coding level decreases.

Although we argue that behavior is a key indicator of people's internal cognitive and emotional states, specifying exactly what those behaviors indicate remains a challenge for behavioral researchers.

Using Behavior in our Social and Personality Psychology Theories

We suggested earlier that it may be difficult for social and personality researchers to generate interesting and meaningful research questions that center around behavior. Part of this is due to the field's current focus on internal states – cognition, emotions, and motivations. Part of this is due to the increasing emphasis on multi-study papers in our top journals. Given the often time consuming task of behavioral observation, it would be difficult to accrue the number of publications needed for academic success when packing those publications with multiple studies that include behavioral observation. And finally, the further away from behavior we've gone as a field, the more difficult it has become to conceptualize our variables in behavioral terms.

Many of the currently popular theories in social and personality psychology reflect this internal focus. For example, in Terror Management Theory (Solomon, Greenberg, & Pyszczynski, 1991), the mortality salience hypothesis states that reminders of one's mortality will increase the needs of individuals to value their own cultural worldview and self-esteem. This may result in behavioral changes, but more typically, these motivational changes are indexed by outcomes such as increased stereotypical thinking and biased intergroup evaluation. And it is those internal changes that are assessed and reported in research publications.

The currently popular theory of persuasion in attitudes research, the Elaboration Likelihood Model (Petty & Wegener, 1999), similarly focuses on the cognitive variable of elaborative thought processing as the mediator between source messages and attitude change. The impact of attitude change on actual behavior is not addressed, although theoretically much is known regarding when attitudes predict behavior.

Again, this is not to say that these inner states are not important. It is certainly important to predict and understand another person's emotional or cognitive reactions to stimuli. We are only saying that this is not the whole picture when trying to understand people. Sometimes we are interested in internal processes in and of themselves. But at other times, these internal processes are meant as precursors to action, and it is important that the action itself be included in the assessment picture as well.

There has been a recent rise in theories and research programs that emphasize behavior. For example, Gollwitzer's work on implementation intentions (Gollwitzer, 1999; 2010) includes both the assessment of internal cognitions – a specific cognitive plan for carrying out an activity – as well as an assessment of the activity itself. With a focus on behavior as the outcome, Gollwitzer and colleagues have shown how different forms of intentions affect the occurrence of an actual behavior (e.g., Gawrilow & Gollwitzer, 2008; Gollwitzer & Sheeran, 2006). Chartrand and Bargh's (1999; Bargh & Morsella, 2010) work on behavioral mimicry also has a strong behavioral component and offers interesting evidence that the environment can automatically determine behavior. And research on interaction synchrony and rapport (Tickle-Degnan, 2006) demonstrates that behavioral coordination can have a profound impact on relational outcomes.

It is also worth noting that social psychological and personality research featuring actual behaviors are more likely to have impact on real-world problems as well as on policymakers (including those who fund our less behavior-oriented research). For example, research on stereotype threat has helped to identify what interventions can be undertaken to reduce academic performance deficits in those students who are the

target of negative group-based stereotypes. Actual behavioral differences (“performance deficits”) between students are tangible and more easily identified as problematic by educators and policymakers alike. Put another way, behavior attracts attention.

It is not that we want to strip the assessment of internal processes from our research, but rather we want to complement those assessments with assessments of external behavioral processes. We believe it would be a mistake to ignore the complexities of the actual enactment of behavior when building a comprehensive understanding of human interaction.

Concluding Thoughts

In the chapter, we have attempted to reintroduce behavior and behavior assessment to personality and social psychology. Although early research in our fields was characterized by studies that included observations of behavior – and sometimes quite dramatic behavior – this cannot be said to be true for research conducted at least for the past twenty-five years. Social and personality psychology attempts to understand how the thoughts, feelings, *and behaviors* of individuals are influenced by individual characteristics and by other people. It is time to more completely embrace this definition by grounding our assessments whenever appropriate in behavioral observation.

References

- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441.
- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. University of Vermont, Department of Psychiatry, Burlington, VT.
- Agnew, C. R. (1999). Power over interdependent behavior within the dyad: Who decides what a couple does? In L. J. Severy & W. B. Miller (Eds.), *Advances in population: Psychosocial perspectives, Volume 3* (pp. 163-188). London: Jessica Kingsley Publishers.
- Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.) (2010). *Then a miracle occurs: Focusing on behavior in social psychological theory and research*. New York: Oxford University Press.
- Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (2010). Behavior and miracles. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 3-11). New York: Oxford University Press.
- Agnew, C. R., & Kelly, J. R. (2010). Behavior between people: Emphasizing the “act” in interaction. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 275-282). New York: Oxford University Press.
- Ajzen, I. (1991). The theory of planned behavior. Special Issue: Theories of cognitive self-regulation. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888-918
- Aron, A. (2010). Behavior, the brain, and the social psychology of close relationships. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 283-298). New York: Oxford University Press.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, *29*, 819-833.
- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, *17*, 102-141.
- Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. K. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 138-159). New York: Cambridge University Press.
- Bales, R. F. (1950). *Interaction process analysis; a method for the study of small groups*. Oxford, UK: Addison-Wesley.
- Bales, R.F. & Cohen, S.P. (1979). *SYMLOG: A system for the Multiple Level Observation of Groups*. New York: Macmillan Publishing.
- Bargh, J. A., & Morsella, E. (2010). Unconscious behavioral guidance systems. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 89-118). New York: Oxford University Press.

- Batson, C. D. (1998). Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey, (Eds.), *The handbook of social psychology*. New York: McGraw Hill.
- Baumeister, R.F., Vohs, K.D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2, 396-403.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 6. NY: Academic Press.
- Bernieri, F. J., Reznick, J. S., Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, 54, 243-253.
- Bolger, N., Stadler, G., Paprocki, C., & DeLongis, A. (2010). Grounding social psychology in behavior in daily life: The case of conflict and distress in couples. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 368-390). New York: Oxford University Press.
- Buss, A. (1961). *The psychology of aggression*, New York: Wiley.
- Carney, D. R., Jost, J. T., Gosling, S. D., & Potter, J. (2008). The secret lives of liberals and conservatives: Personality Profiles, interaction styles, and the things they leave behind. *Political Psychology*, 29, 807-840.
- Chapple, E. D. (1970). *Culture and biological man: Explorations in behavioral anthropology*. New York: Holt, Rinehart, & Winston.

- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893-910.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*, 1015-1026.
- Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.
- Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development*, *67*, 2317-2327.
- Dabbs, J. M., & Ruback, R. B. (1987). Dimensions of group process: Amount and structure of vocal interaction. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 123-169). New York: Academic Press.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *28*, 377-383.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *The facial action coding system* (2nd ed.). Salt Lake City, UT: Research Nexus eBook.
- Fagot, B. T. & Hagan, R. (1985). Aggression in toddlers: Responses to the assertive acts of boys and girls. *Sex Roles*, *12*, 341-351.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Foushee, H. C. (1984). Dyads and triads at 35,000 feet. *American Psychologist*, *39*,

885-893.

Foushee, H. C., Lauber, J. K., Baetge, M. M., & Acomb, D. B. (1986). *Crew factors in flight operations: III. The operational significance of exposure to short-haul air transport operations*. Moffet Field, CA: NASA Ames Research Center, (NASA Technical Memorandum 88322).

Funder, D. C. (2006). Toward a resolution of the personality triad: Persons, situations and behaviors. *Journal of Research in Personality, 40*, 21-34.

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-Sort: A tool for the description of social behavior. *Journal of Personality, 68*, 451-489.

Furr, R. M., & Funder, D. C. (2007). Behavioral observation. In Robins, Richard W (Ed); Fraley, R. Chris (Ed); Krueger, Robert F (Ed). (2007). *Handbook of research methods in personality psychology*. (pp. 273-291). xiii, 719 pp. New York, NY, US: Guilford Press.

Furr, R. M., Wagerman, S. A., & Funder, D. C. (2010). Personality as manifest in behavior: Direct behavioral observation using the Revised Riverside Behavioral Q-Sort (RBQ-3.0). In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 186-204). New York: Oxford University Press.

Gaddis, B, Connelly, S., & Mumford, M. D. (2004). Failure feedback as an affective event: Influences of leader affect on subordinate attitudes and performance. *The Leadership Quarterly, 15*, 663–686.

Gawrilow, C., & Gollwitzer, P. M. (2008). Implementation intentions facilitate response inhibition in children with ADHD. *Cognitive Therapy and Research, 32*, 261-280.

- Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology, 94*, 91-107.
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*, 493-503.
- Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology, 38*, 69-119.
- Gollwitzer, P. M., Wieber, F., Myers, A. L., & McCrea, S. M. (2010). How to maximize implementation intention effects. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 275-282). New York: Oxford University Press.
- Gosling, S. D., Craik, K. H., Martin, N. R., & Pryor, M. R. (2005). The Personal Living Space Cue Inventory: An analysis and evaluation. *Environment and Behavior, 37*, 683-705.
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Judgments of personality based on offices and bedrooms. *Journal of Personality and Social Psychology, 82*, 379–398.
- Hartshorne, H., & May, M. A. (1928). *Studies in the Nature of Character*. New York: Macmillan.
- Holmes, J. G., & Cavallo, J. V. (2010) The atlas of interpersonal situations: A theory-driven approach to behavioral signatures. In Agnew, C. R., Carlston, D. E.,

- Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 321-341). New York: Oxford University Press.
- Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.
- Kelly, J. R. (2000). Interaction process analysis in task performing groups. In A. P. Beck and C. M. Lewis (Eds.), *Process in Therapeutic Groups: A Handbook of Systems of Analysis*. Washington, DC: APA Books.
- LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces*, 13, 230-237.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815-852.
- Loftus, E. F., & Cahill, L. (2007). Memory Distortion: From Misinformation to Rich False Memory. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 413-425). NY: Psychology Press.
- Lucas, P. J. J., Noldus, R. J. H., Trienes, A. H. M., Hendriksen, H. J., & Roland G. J. (2000). The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, & Computers*, 32, 197-206.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.
- Mehl, M. R. (2009). Naturalistic observations of daily behavior in personality psychology. *European Journal of Personality*, 23, 414-416.

- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*, 862-877.
- Mehl, M. R., Vazie, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317*, 82.
- Milgram, S. (1975). *Obedience to authority*. New York: Harper & Row.
- Mischel, W. (1968). *Personality and assessment*. NY: Wiley.
- Moreland, R. L., Fetterman, J. D., Flagg, J. J., & Swanenburg, K. L. (2010). Behavioral assessment practices among social psychologists who study small groups. In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 28-56). New York: Oxford University Press.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 2*, 436-50.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.
- Northouse, P. G. (2007). *Leadership Theory & Practice* (4th Ed.). London: Sage.
- Ohio State Leadership Studies. (1964). Retrieved from <http://fisher.osu.edu/offices/fiscal/lbdq>.
- Ostrov, J. M., & Keating, C. F. (2004). Gender Differences in Preschool Aggression During Free Play and Structured Interactions: An Observational Study. *Social Development, 13*, 255-277.

- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver, & L. S. Wrightsman (Eds.). *Measures of personality and social psychological attitudes* (pp. 17-60). San Diego: Academic Press.
- Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 227-246). New York: Oxford University Press.
- Pellegrini, A. D. (2001). Practitioner review: The role of direct observation in the assessment of young children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 861–869.
- Petty, R. E., & Wegener, D. T. (1999). The Elaboration Likelihood Model: Current Status and Controversies. In S. Chaiken & Y. Trope (eds.), *Dual Process Theories in Social Psychology* (pp. 41–72). New York: Guilford Press.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *3*, 377-386.
- Sande, G. N., Goethals, G. R., & Radloff, C. E. (1988). Perceiving one's own traits and others': The multifaceted self. *Journal of Personality and Social Psychology*, *54*, 13-20.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961): *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University of Oklahoma Book Exchange.
- Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 18. NY: Academic Press.

- Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). Terror management theory of self-esteem. In C. R. Snyder & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective* (pp. 21-40). Elmsford, NY: Pergamon Press.
- Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: The relative accuracy of self vs. others for the prediction of behavior and emotion. *Journal of Personality, 68*, 837-867.
- Tickle-Degnen, L. (2006). Nonverbal behavior and its functions in the ecosystem of rapport. In M. L. Patterson & V. Manusov (Eds.), *Handbook of nonverbal communication*. Thousand Oaks, CA: Sage.
- Utz, S. (2010). Using automated "field notes" to observe the behavior of online subjects. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 91-108). Washington, DC: American Psychological Association.
- Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology, 45*, 1028-1031.
- Wallace, P. (2001). *The psychology of the Internet*. New York, NY: Cambridge University Press.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove J. B. (1981). *Nonreactive measures in the social sciences*. Boston, MA: Houghton Mifflin Company.

- Weingart, L. R. (1997). How did they do that: The ways and means of studying group processes. *Research in Organizational Behavior* 19, 189-239.
- Weingart, L. R., Hyder, E. B., & Prietula, M. J. (1996). Knowledge matters: The effect of tactical descriptions on negotiation behavior and outcome. *Journal of Personality and Social Psychology*, 70, 1205-1217.
- Wicker, A. W. (1969). Attitudes vs. actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 25, 41-78.
- Williams, K. D., & Sommer, K. L. (1997). Social ostracism by coworkers: Does rejection lead to loafing or compensation? *Personality and Social Psychology Bulletin*, 23, 693-706.
- Verplanken, B. (2010). Habit: From overt action to mental events. In Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 89-118). New York: Oxford University Press.
- von Eye, A., & Mun, E. Y. (2005). *Analyzing rater agreement: Manifest variable methods*. Mahwah, NJ: Lawrence Erlbaum.

Table 1. Important Dimensions of Behavior

1. Intentional	←-----→	Unintentional
2. Conscious	←-----→	Nonconscious
3. Independent	←-----→	Interdependent
4. Discrete	←-----→	Continuous
5. Distinct	←-----→	Embedded

Table 2. Variability in Behavior that Affects Measurement

1. Novel ←-----→ Habitual
2. Low Frequency ←-----→ High Frequency
3. Desirable ←-----→ Undesirable

Figure 1. Stages of Behavioral Observation

