

Chapter 3

Social Psychological Methods Outside the Laboratory

HARRY T. REIS AND SAMUEL D. GOSLING

When Kurt Lewin ushered in the modern era of experimental social psychology, he did so with the strong belief that the scientific psychology of the time seemed to be trying “increasingly to stay away from a too close relation to life” (1951, p. 169). Lewin primarily intended to keep experimental social psychology close to life by urging researchers to maintain an active interest in applications of theory to social problems, but he also felt that, beyond research with experimentally created laboratory groups, the field

shall have also to develop research techniques that will permit us to do real experiments within existing “natural” social groups. In my opinion, the practical and theoretical importance of these types of experiments is of the first magnitude. (1951, pp. 164–165)

By this Lewin meant that social psychological research needed to keep its theoretical feet firmly grounded in real-world contexts, problems, and social relations.

In the more than half-century of research and theorizing that followed, social psychology’s remarkable progress has derived in large measure from laboratory research. For example, Sears (1986) reported that 78% of the social-psychological research published in 1985 in the field’s top journals was conducted in the laboratory. Rozin (2001) similarly concluded that nearly all of the articles published in the first two sections of volume 66 (1994) of the *Journal of Personality and Social Psychology* (JPSP) were situated in the laboratory or used questionnaires. No doubt this emphasis reflects the many benefits of laboratory (typically, although not exclusively experimental) research, including experimental control over variables, contexts and procedures, which allows researchers to control extraneous

influences and differentiate causal mechanisms from one another (Smith, 2000), and easy access to undergraduate samples. These advantages were a great part of the reason why social psychology, which had been more non-experimental than experimental in its early days, evolved into a predominantly experimental science during the 1930s and 1940s (House, 1977; Jones, 1985), a considerable and enduring legacy.

But these advantages may also have a cost, in terms of increasing distance from Lewin’s “close relation to life.” Laboratory settings by definition remove research participants from their natural contexts and place them in an artificial environment in which nearly all aspects of the setting, including physical features, goals, other persons involved, and even the possibility of getting up and doing something else, are determined by an external entity (i.e., the experimenter). Natural habitats, in contrast, are marked by far greater diversity and clutter of the physical and social environment, the necessity of choosing for oneself what task to pursue and how to engage it, and the option of changing settings and tasks. Ironically, social-psychological research has provided ample testimony of the importance of context for understanding behavior.

The good news is that social psychology can have it both ways. As is discussed below, researchers have come to realize that validity is not an “either-or” proposition but rather the result of complementary methods targeting the same theories, processes, and concepts. Just as social psychologists have used stagecraft to import some of the richness of natural settings into the laboratory, recent methodological advances have made possible with non-laboratory methods some of the same precision and control that

For their helpful comments on earlier drafts of this article, we gratefully thank Matthias Mehl, Peter Caprariello, Michael Maniaci, Shannon Smith, and the editors.

Direct correspondence to Harry T. Reis, Ph.D., Department of Clinical and Social Sciences in Psychology, Box 270266, University of Rochester, Rochester, New York 14627; voice: (585) 275-8697; fax: (585) 273-1100; e-mail: reis@psych.rochester.edu.

heretofore was possible only in the laboratory. Moreover, many of these advances allow non-laboratory research to ask more complex questions or to obtain far more detailed responses than the typical laboratory experiment. As a result, non-laboratory methods represent a far more powerful tool for social psychological research and theory than they have previously. Adding them to a research program may also make the results of research more interesting and relevant, as Cialdini (2009) suggests.

The distinction between laboratory and non-laboratory research is sometimes conflated with sampling. Although undergraduate and non-undergraduate samples are studied in both kinds of settings, in actuality the vast majority of laboratory studies rely on undergraduate samples, whereas non-laboratory studies are more likely to use non-student, adult samples. Eighty-three percent of the studies in Sears's (1986) review used samples composed of students. Reviews of the 1988 volume of *JPSP*, the 1996 volume of the *Personality and Social Psychology Bulletin*, and the 2002 volume of *JPSP* put these estimates at 80%, 85%, and 85%, respectively (Gosling, Vazire, Srivastava, & John, 2004; Sherman, Buddie, Dragan, End, & Finney, 1999; West, Newsom, & Fenaughty, 1992). Another study reported that 91.9% of studies of prejudice and stigma published in the field's top three journals from 1990 to 2005 relied on undergraduate samples, and even in two expressly applied journals (*Journal of Applied Social Psychology*, *Basic and Applied Social Psychology*), 73.6% of studies were based on research with undergraduates (Henry, 2008). Laboratory studies use undergraduate samples because it is difficult and expensive to recruit nonstudent participants to come to the lab. With non-laboratory studies, researchers usually have little reason to prioritize nonstudent samples.

This chapter reviews some of the more important, popular, and timely methods for conducting social psychological research outside of the laboratory. The chapter begins with a review of the purpose of non-laboratory methods, emphasizing how they have been used in social psychology, as well as the kinds of insights that they can and cannot provide. Included in this section is a review of how laboratory and non-laboratory methods complement each other in a research program. We then describe in some detail five methods that have become influential tools in social psychology and give every indication of continued value: field experiments, Internet methods, diary methods, ambulatory monitoring, and trace measures. The chapter concludes with a brief commentary on the future of non-laboratory methods in social psychology.

We do not review two broad and common classes of non-laboratory methods, survey research and observational methods, for space reasons and because excellent

discussions of these methods are available elsewhere. Readers interesting in learning more about survey methodology may consult the chapter by Schwarz, Groves, and Schumann in the fourth edition of this *Handbook* (1998), Krosnick and Fabrigar (in press), Groves et al. (2004), or Visser, Krosnick, and Lavrakas (2000). Fuller description of observational methods (which are applied both in the laboratory and in non-laboratory settings such as work sites, homes, and schools) may be found in Weick's (1985) chapter in the third edition of this *Handbook*, or in Bakeman (2000), Bakeman and Gottman (1997), Kerig and Lindahl (2001), and McGrath and Altermatt (2000). Other non-laboratory methods used by social psychologists that we do not discuss include archival methods (Simonton, 2003; Webb, Campbell, Schwartz, & Sechrest, 2000), computer simulations (Hastie & Stasser, 2000), interviews (Bartholomew, Henderson, & Marcia, 2000), and participant observation in the field.

WHAT IS MEANT BY NON-LABORATORY RESEARCH?

We are tempted to define the term *non-laboratory research* as all research conducted elsewhere than in a laboratory suite, room, or cubicle. Laboratories are spaces specially equipped for research that permit experimenters to control nearly all facets of the participant's experience, including the physical (e.g., ambient sound and temperature, furniture, visual cues) and social environment (e.g., other persons), as well as the possibility of distraction by external circumstances (e.g., cell phones). Conducting non-laboratory research necessarily involves sacrificing this high level of control over extraneous factors for the benefits discussed below. Researchers often design non-laboratory studies to observe social-psychological phenomena in their natural context, reflecting the belief that the setting in which a behavior occurs must be a fundamental part of any theoretical account of that behavior (Weick, 1985). (This belief is of course entirely consistent with the rationale for laboratory research, because settings would not need to be controlled if they were not influential.) In contrast, the laboratory setting is likely to engender certain expectations and scripts (e.g., serious purpose, scientific legitimacy, the possibility of deception, the importance of attentiveness), which may affect behavior (Shulman & Berman, 1975). Non-laboratory research also tends to constrain participant behavior less, in the sense that the setting offers many more alternative activities (e.g., participants can choose what to do, when, where, and with whom) and distractions, so that self-direction and spontaneous selection among activities is greater. In a laboratory study, participants usually can

do little else but complete the tasks assigned to them by researchers.

Laboratory and non-laboratory settings differ in various ways, some of them more influential than others. Administering a questionnaire in a classroom versus a laboratory cubicle may not make much difference, whereas conducting a field experiment on the effects of affectionate smiles on attraction at a social mixer versus a laboratory room may matter more. Non-laboratory and laboratory contexts differ in three general ways: the physical environment, the goals likely to be activated by the setting and their correspondence with the behaviors being studied, and the degree to which the setting is natural and appropriate for the research question. Of these, we see the latter two factors as more significant for social-psychological research. That is, because behavior reflects personal goals and concerns, and is embedded in naturally occurring contexts, non-laboratory research can complement laboratory studies best when it highlights such influences.

It is important to note that the setting in which a study is conducted is independent of whether a study is experimental or non-experimental (see Figure 3.1). Studies conducted outside of the laboratory can possess all of the features of a true experiment—random assignment to conditions, manipulation of the treatment conditions—as in the case of field experiments and randomized clinical trials (Wilson et al., this *Handbook*), just as studies conducted in a laboratory space can have an experimental or correlational design. Some studies include both laboratory and non-laboratory components, such as when measurements obtained in the laboratory are used to help explain behaviors observed in non-laboratory settings. Also, although non-laboratory research may possess less of the tight control over setting and procedure that is typically associated with laboratory research, systematic, carefully designed methods are still essential.

Their relative infrequency notwithstanding, non-laboratory studies have played an important role in social

psychology, both historically and in contemporary research. A few examples may illustrate this role, as well as highlight the diversity of methods included in this general category.

Field studies (including experiments, quasi-experiments, and nonexperimental designs) include the famous Robbers Cave research, conducted in 1954, in which observations of early adolescent boys attending a summer camp led to findings about ingroup cooperation and outgroup competition that spawned one of social psychology's most enduring research areas, intergroup conflict (Sherif, Harvey, White, Hood, & Sherif, 1961). The development of cognitive dissonance theory was influenced in an important way by *When Prophecy Fails*, a field study in which researchers infiltrated a prophetic group of doom-sayers predicting the end of the world (Festinger, Riecken, & Schachter, 1956). Many important studies of bystander intervention in the 1960s and 1970s took place in natural settings, such as grocery stores, streets, the New York City subways, and Jones Beach (Moriarty, 1975). Important studies examining if, when, and how intergroup contact reduces prejudice and discrimination have been conducted with actual conflicting groups (Amir, 1969), and real-world classrooms have been used to study the effects of cooperative learning structures (the so-called *Jigsaw Classroom*) on intergroup relations and academic achievement (Aronson, 2004; Johnson, Johnson, & Smith, 2007). Some of the earliest studies of self-disclosure processes were conducted with Navy recruits in boot camp training for service on submarines and other isolated yet intensely interactive settings (e.g., Altman & Haythorn, 1965). Pioneering studies of the acquaintance process observed the development (and non-development) of friendships among new students at Bennington College and the University of Michigan (Newcomb, 1961). More contemporary examples of field research in social psychology include studies of personal living and working spaces (Gosling, Ko, Mannarelli, & Morris, 2004), investigations of attachment processes within the Israeli military (Davidovitz, Mikulincer, Shaver, Izsak, & Popper, 2007), and Sherman and Kim's (2005) studies of self-affirmation among student members of sports teams. The Internet has also created many new possibilities for research.

Ambulatory monitoring (including diary methods) has grown rapidly in recent years, no doubt due to technological advances that make such procedures more accessible and cost-effective while providing better, more detailed information. Among the most popular of these methods are *Experience Sampling* (Hektner, Schmidt, & Csikszentmihalyi, 2007) and *Ecological Momentary Assessment* (Stone & Shiffman, 1994), both defined later in this chapter, which have been used extensively to study affect, cognition,

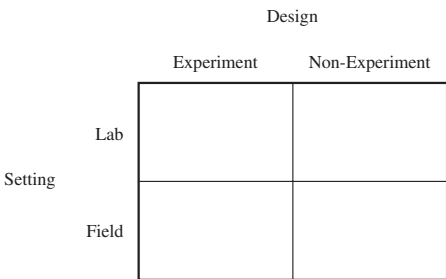


Figure 3.1 Designs and Settings Are Orthogonal.

health symptoms, health-related behavior, social interaction, and activity in everyday life. Ambulatory assessment procedures have also been used in social psychological research to collect random samples of the acoustic environment (Pennebaker, Mehl, & Niederhoffer, 2003); to obtain detailed reports of physiological states, particularly heart rate variability and other cardiovascular measures (Hawkey, Bursleson, Berntson, & Cacioppo, 2003), as they relate to what the person is doing or experiencing; to characterize sleep (Ajilore, Stickgold, Rittenhouse, & Hobson, 1995); and to quantify person-to-person proximity for social network analyses (Pentland, 2007).

WHY STUDY SOCIAL PSYCHOLOGICAL PROCESSES OUTSIDE THE LABORATORY?

Debate over the relative priority that should be given to internal and external validity is not new. Many commentators have bemoaned the seeming low priority given to generalizability (e.g., Helmreich, 1975; McGuire, 1967; Ring, 1967; Silverman, 1971; see Henry, 2008, for a more recent version). Various replies have been provided, the most commonly cited of which argue that the purpose of laboratory experiments is to evaluate theories, regardless of the applicability of those theories—in other words, to determine “what can happen” as opposed to “what does happen” (e.g., Aronson, Wilson, & Brewer, 1998; Berkowitz & Donnerstein, 1982; Mook, 1983). Perhaps reflecting the effectiveness of these replies, social psychologists are usually taught that internal validity has higher priority than external validity—that it is more important to be certain about concluding that an independent variable is the true source of changes in a dependent variable than it is to know that research findings can be generalized to other settings and samples. Too often, however, in our opinion, the lesser priority of external validity is confused with low priority, which fosters a certain irony. Social psychology generally seeks principles to describe social behavior that hold across persons, settings, and (sometimes) cultures (Cook & Groom, 2004). How do we know this to be so without research that establishes the point? Non-laboratory methods are well suited to demonstrating external validity.

As most discussions of methodology point out, no single study can maximize all types of validity (e.g., Brewer, 2000; Smith & Mackie, 2000). All methods have their advantages and drawbacks, which is why methodological pluralism—using multiple and varied paradigms, operations, and measures to triangulate on the same concepts—has long been advocated as a feature of research programs (Campbell, 1957; Campbell & Fiske, 1959), if more in principle than in practice. There is little reason to doubt that

laboratory settings tend to be preferable for conducting the most carefully controlled studies, because manipulations can be crafted to precisely test some theoretical principle, controlling for the “noise” of the real world and ruling out alternative explanations and potential artifacts (even those that may be confounded with the key independent variable in natural experience). Nevertheless, the value of non-laboratory research goes well beyond showing that the same processes are also evident in the real world. As Brewer notes, “the kind of systematic, programmatic research that accompanies the search for external validity inevitably contributes to the refinement and elaboration of theory as well” (2000, p. 13).

Brewer (2000) described the three Rs of how external validity contributes to the development of theory and knowledge:

1. *Robustness*, or whether a finding is replicated in different settings, with different samples, or in different historical or cultural circumstances. Although researchers sometimes couch replications of this sort in checklist terms (“yes it did” or “no it didn’t”), it is more informative to think about replications in terms of their ability to identify boundary conditions for an effect and other moderating variables, which in turn may contribute to fuller understanding of the scope, context, and mechanism for a phenomenon. For example, most social psychologists believe that behavior is a function of Person \times Environment interactions (Funder, 2006), and such interactions are more likely to be revealed in studies with heterogeneous populations. Similarly, ever since Barker (1968), most social psychologists have believed that settings affect behavior, yet in laboratory studies, although setting variables may be controlled (perhaps as part of “lab lore”), they are not systematically investigated. The situational variable held constant in one program of research may be the focal variable of another research program. Replications, in other words, help identify moderator variables that are essential to the full specification of a theory and its component processes. Because laboratory studies typically isolate the variables in question from influence by settings, individual differences, and other contextual factors in order to identify cause-and-effect associations, they tend to privilege main effects (Cook & Groom, 2004).
2. *Representativeness*, or do the conditions or processes actually occur in the real world? This differs from robustness because an effect might be highly replicable, but unlike anything that occurs in normal circumstances. Brunswik (1956) pointed out the importance of representativeness, in noting that generalizability to



the real-world depended on random sampling of both participants and contexts. Nonetheless, the biological and physical sciences commonly use unrepresentative conditions to test theory (e.g., the behavior of electrons in a vacuum may illuminate a proposed mechanism), and they may have similar value for social-psychological theory (Petty & Cacioppo, 1996). For example, examining the effects of mere exposure with variably mixed combinations of familiar and unfamiliar stimuli may not resemble circumstances that naturally occur, but they allow researchers to compare explanations based on fluency and repetitiveness (Dechêne, Stahl, Hansen, & Wänke, under review). Nonetheless, understanding when and how processes apply to natural social behavior necessarily provides a foundation for theory development, just as descriptive taxonomies of species provide a foundation for biological research (Kelley, 1992). Furthermore, identification of the circumstances under which a phenomenon occurs in the real world may suggest important clues about covariates, mechanisms, and limiting conditions (e.g., as has been shown in repeated efforts to apply the contact hypothesis to actual intergroup conflicts [Pettigrew & Tropp, 2006]). Representativeness is also important for translations and application of basic research.

3. *Relevance*, or can the findings be used to modify behavior in the real world? Of course not all research (laboratory or non-laboratory) is intended to test intervention-related hypotheses, but to the extent that theories can be used to modify behavior, their theoretical basis is strengthened. This principle underlies Lewin's (1951) belief in the value of "action research" for theory development, as well as the more general claim that psychological theories are useful if they can be used to predict and control behavior. Because non-laboratory applications do not isolate the effects of a given manipulation from the simultaneous effects of other processes in the natural environment, they help identify the relative strength of a given manipulation in context, as well as its sensitivity to interference by moderating variables. (It is easy to imagine circumstances in which a manipulation might produce effects of considerable effect size under the tightly controlled conditions of the laboratory, yet be ineffectual in the real world.) Haslam and McGarty (2004) suggest an inverse relationship between relevance and sensitivity: The more relevant a given issue to participants, the less sensitive (i.e., modifiable) their behavior may be. For example, in most cases it would be easier to modify lawn care behavior than sexual behavior, even though the same general theory may apply.

A somewhat different way of conceptualizing the relative advantage of non-laboratory research concerns the issue of closeness to real-world concerns (closely related to, but not the same as, the distinction between *mundane realism*, or, the extent to which the events of an experiment resemble real-world events, and *experimental realism*, or, the extent to which experimental events are involving; see Wilson et al., this *Handbook*). Weick (1985) posed a series of intriguing questions about which situations get "closer" to the human condition: A study of how one tells a newly acquainted stranger in the laboratory that she is about to receive a mildly painful electric shock or a study of how a coroner announces death to next of kin. Or, anticipation of putting one's hand in a bucket of ice water in a controlled laboratory room or learning how to work on high steel in a 21-story building. Distance, Weick argued, may encourage ambiguity and detachment from the motives, wishes, fears, and concerns that drive behavior in the real world.

To be sure, laboratory studies can be intensely involving, but often they are not (Baumeister, Vohs, & Funder, 2007), especially in light of the restrictions that Research Ethics Boards increasingly demand, which make it difficult for researchers to engage participants in a way that activates strong personal involvement. If the setting is chosen properly, such involvement is readily accessible in non-laboratory studies—for example, the same undergraduate student who is only mildly concerned about having performed poorly on a laboratory task of mental arithmetic may be substantially more engaged in the outcome of her calculus midterm examination. Similarly, recent speed-dating research has yielded results that differ from more traditional laboratory studies of initial romantic interactions (Finkel & Eastwick, 2008). Non-laboratory studies, in other words, may bring research questions "closer" to involving, personally meaningful motives, defenses, affects, and thought processes.

Just how effectively non-laboratory studies accomplish these goals depends, of course, on how the research is designed and conducted. Non-laboratory studies need to be systematic, coherent, and controlled for the impact of errors and artifacts; a flawed field study contributes no more than a poorly designed laboratory experiment. No individual study can simultaneously minimize all threats to internal validity by experimental control, nor all possible limits on generalizability by going outside the laboratory. Validity, in the broadest sense, depends on matching protocols, designs, and methods to questions, so that, across a program of research, all reasonable alternative explanations are ruled out and boundary conditions are established. Thus, as with laboratory research, the ultimate rationale for conducting non-laboratory research is to advance the depth, accuracy, and usefulness of social-psychological knowledge.

FIELD EXPERIMENTS

As mentioned earlier, experiments, quasi-experiments, and non-experimental (correlational) designs can be enacted in field settings. The principles that distinguish these designs from one another are the same, regardless of whether the research is conducted in the field or in the laboratory; consequently, readers are referred to the chapter by Wilson et al. (this *Handbook*) for discussion of the basic principles of experimentation. It bears noting that a great deal of field research is non-experimental in nature—for example, simple observational studies in which the behavior of persons in natural habitats is observed. We do not discuss those methods here; for simplicity, we use the term “field experiments” to refer to field experiments and field quasi-experiments, although we intend no conceptual confusion between the terms.

Researchers conduct field experiments for several reasons. The desire to maximize external validity is cardinal among them, as discussed earlier. Another reason is the desire to observe phenomena in their natural contexts, without controlling for other influences, so that processes can be studied within the full circumstances in which they are most likely to occur (Reis, 1983). This principle refers to whether the conditions in research are representative of the typical conditions in which that effect commonly occurs.¹ A third advantage of field experiments is that most often, participants are not aware of being in a psychology experiment, thereby minimizing *demand characteristics* (cues that suggest to research participants the behaviors that researchers expect of them), suspicion, and other reactive effects that may occur in the laboratory context. A final reason is that some researchers simply find field settings “more interesting” (Salovey & Williams-Piehot, 2004), although, we hasten to add, for other researchers the same sentiment may apply to laboratory research.

Consider a study conducted by Bushman (1988). In this study, a female confederate approached pedestrians and instructed them to give change to an accomplice standing next to an expired parking meter. To investigate the effects of perceived authority on compliance, the confederate wore one of three outfits: a uniform, business clothes (to imply status but not authority), or sloppy clothes that made her appear to be a panhandler. The uniform condition induced greater compliance than the other two conditions, which did not differ significantly from each other. This setting is a natural one for this kind of request and for both

the independent (attire) and dependent (giving a coin to the accomplice) variables. It is unlikely that participants suspected that they were in an experiment or that their response to the attire was under scrutiny. Had the same experiment been conducted in the laboratory, participants might well have been more attentive to these possibilities. (Of course, in the laboratory, it would be easier to manipulate perceived authority in a way that kept confederates unaware of conditions, so that their behavior could not have varied systematically across conditions.) Additionally, participants cannot walk away muttering “sorry” in the laboratory, as they can in real life.

The inability to gain control over extraneous circumstances that might have influenced the findings is the chief disadvantage of field experiments. In Bushman’s simple experiment, these seem unlikely. But consider a field experiment conducted by Josephson (1987), in which second- and third-grade boys were frustrated before or after watching violent or nonviolent television programs in school, then observed playing floor hockey with other children. Because of random assignment to conditions, we can be confident that the conditions were responsible for observed differences in aggressiveness but various uncontrolled factors may also have been influential: How closely did the boys attend to the programs? Did the adults present respond to the boys in ways that facilitated or inhibited aggression? Were there cues in the school that influenced their responses? Did interaction among the children alter their responses? Questions of this sort are central to identifying the mechanism responsible for an effect, and it is likely that these factors could have been controlled better in a laboratory experiment.

Researchers more commonly conduct quasi-experiments in field than in laboratory settings, and because participants in quasi-experiments are not randomly assigned to conditions, threats to internal validity tend to be greater. For example, had the boys in Josephson’s study not been randomly assigned to conditions, but instead had one classroom been assigned to watch violent programs and another classroom to watch nonviolent programs, other factors (e.g., pre-existing differences between the classrooms, other classroom events during the study interval) might plausibly have caused the observed differences. For this reason, quasi-experiments involve pre-manipulation and post-manipulation assessments, and typically include as many other design elements as possible to address these threats to internal validity (Cook & Campbell, 1979; West, Biesanz & Pitts, 2000).

Field experiments often alter the typical balance between mundane and experimental realism. As originally defined by Aronson and Carlsmith (1968), mundane realism is high when a research protocol resembles events likely to occur

¹ Although this is sometimes referred to as ecological validity, Hammond (1998) points out that this term represents a misleading application of what Brunswik, who originated the term, meant.

in normal activity. Experimental realism, in contrast, is high when participants find research involving and engrossing so that they are interested, attentive, and motivated to take their task seriously. Laboratory research generally puts higher priority on experimental than mundane realism, for reasons explained by Wilson et al. (this *Handbook*). Field research almost by definition maximizes mundane realism, because participants are encountered in their normal activity, although perhaps ironically, experimental realism may not be high. For example, persuasive messages or a request for help delivered casually and ineffectually by a stranger in a coffee shop may be dismissed in a cursory manner, with little or no thought or concern. Or, distracted passers-by may not even notice events staged to take place on a busy street corner, in which many stimuli compete for attention. Researchers should not assume experimental realism in field settings; establishing it requires as much (and perhaps more) care as it does in the laboratory. Of course, some studies are designed to examine processes that operate with minimal engagement (e.g., automaticity), and in this circumstance low experimental realism may be appropriate.

Sometimes, significant real-world events lead researchers into the field, either because that event creates a natural manipulation for what has been studied in the lab (e.g., Zucker, Manosevitz, and Lanyon's 1968 study of affiliation and birth order during the November 1965 New York City blackout) or because the event is so inherently compelling that a research response is called for (e.g., responses to 9/11; Silver, 2004). Such studies most commonly survey responses to the events, but quasi-experiments and experiments are also feasible. For example, one group of researchers conducted linguistic analyses of data collected by an online journaling service for two months before and after the 9/11 attacks (Cohn, Mehl, & Pennebaker, 2004). In another example, researchers used archived letters to the editors of local newspapers to study coping responses over time to the Mount St. Helen's volcano eruptions (Pennebaker & Newton, 1983). Pre-data for natural events may also be available fortuitously; in one instance the researchers had been conducting a short-term longitudinal study of falling in love when the 1989 Loma Prieta earthquakes hit the San Francisco Bay Area (Aron, Paris, & Aron, 1995). Events such as these often allow researchers to tell a gripping story, but because it is usually impossible to control key independent variables or to collect pre-event data retrospectively, threats to internal validity may be substantial.

Below we briefly discuss several issues for researchers planning non-laboratory studies to consider.

Sampling and Random Assignment

Field research is often conducted to obtain samples that are more representative than undergraduate samples. This

need not be the case, however. Studies conducted in or near specialized settings (e.g., football stadia, bridal shows, singles' bars, farmers' markets, or on Wall Street or Telegraph Avenue in Berkeley) may also be unrepresentative, in the sense of providing a non-random sample of persons. Aside from the possibility that an effect operates differently in one nonrandom sample than in another, nonrandom samples may possess restricted range on key variables, which can attenuate results and obscure potential moderators (Cohen, Cohen, West, & Aiken, 2003). Moreover, in quasi-experimental and correlational field studies, the factors that lead participants to one or another condition of a study may introduce the possibility of substantial alternative explanations. For example, a study of participants at a Democratic or Republican presidential rally would need to contend with the fact that there are likely many differences between these groups beyond the candidate supported.

It also can be difficult to randomly assign participants to conditions in field settings. Participants might be more unwilling to take part in an effortful, costly, or unpleasant condition of an experiment than in a less effortful, less costly, or more pleasant condition, a potential threat to non-equivalence of groups and hence internal validity (West et al., 2000). This can be particularly vexing for intervention studies, in which demanding interventions (e.g., for smoking cessation) may foster greater attrition in treatment groups than in wait-list control groups. Or sometimes, the lesser degree of control that inheres in field settings may allow participants to undermine random assignment. For example, teachers might be randomly assigned to run some classrooms in a very cold and controlling manner but others in a warmer, more supportive way. Nonetheless, when in the classroom and faced with instructional demands and other distractions, teachers may behave as they see fit, ignoring, misinterpreting, or contradicting the conditions to which they were assigned.² Of course, researchers can and do take steps to monitor and content with these potential problems; our point is that in field experiments, participants may make choices that interfere with well-designed experimental plans.

Finally, field experiments may suffer from unintentional experimenter bias in the selection of participants and their assignment to conditions. In laboratory studies, experimenters typically do not choose participants, and they assign participants to conditions either before arrival or without possible bias (e.g., by a computer program). In field studies, however, experimenters sometimes chose

² Of course, this may also be a factor in laboratory experiments, but because the experimenter has greater control over what transpires, it is less likely.

whom to approach in a public setting (e.g., “the next person walking alone to turn the corner”) or which condition to assign a participant. In principle experimenters have no discretion over these decisions but in practice experimenters are sometimes tempted to skip a potential participant who looks uncooperative or unfriendly, or to assign an unattractive person to a condition that would require less interaction (which would be equally problematic in the lab). It is important to obviate such biases.

Choice of Outcome Measures

Because field studies are often designed so that “the scientist’s intervention is not detectable by the subject and the naturalness of the situation is not violated” (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981, p. 143), they often use unobtrusive, non-reactive behavioral outcome measures. Although this tendency is not inviolate—field studies often rely on self-report, and lab experiments may use unobtrusive measures (see, for example, Ickes’s 1983 *Unstructured Interaction Paradigm*, in which participants’ spontaneous interactions are videotaped without their awareness)—field studies invite researchers to develop and use outcome measures that index the processes under investigation without raising participants’ awareness that researchers are scrutinizing their behavior. Non-experiments have the further goal to avoid altering or modifying participants’ behavior from what they would otherwise do. These settings create a need to balance creativity and relevance (Does the construct actually apply in this setting?) against validity (Does the measure assess the process it purports to assess?) and sensitivity (Does the proposed measure vary systematically and in measurable increments corresponding to the predictor variable?). Because field research tends to involve more variability in settings and samples than laboratory experimentation, measure development may take relatively more time and effort.

A brief and non-representative sampling of measures used in field studies illustrates the kind of creativity that characterizes successful field research. We distinguish *passive observation*—in which data collection exerts no meaningful effect on the behaviors being assessed—from *active observation*—in which participants respond to some sort of circumstance or manipulation created by the experimenter. (This is slightly different than the distinction between *non-reactive* and *reactive* assessment, which refers to whether participants are required to respond to something or whether the data are already available.) Classic examples of passive observation include Triplet’s (1898) observation that bicycle racers tended to race faster when in the presence of other racers than when alone and Cialdini et al.’s (1976) tally of the tendency of students’

to wear school-identifying clothes as a function of football victories and losses, supporting the idea of “basking in reflected glory.” In another example, seating choices of bus commuters in Singapore display ingroup preferences as a function of sex and ethnicity, but not age (Sriram, 2002). And, in a study of Judo participants in the 2004 Athens Olympics (Matsumoto & Willingham, 2006), photographers took action shots at several points. Gold and bronze medal winners were more likely to display Duchenne (spontaneous, genuine) smiles than silver medal winners, supporting the role of counterfactual thinking in emotional experience. Behaviors characteristic of attachment (e.g., clinging, crying, hugging, holding hands) demonstrably occur among adults separating at airports (Fraley & Shaver, 1998). A final example comes from the test of a hypothesis about the role of concealed ovulation in human mating. Professional lap dancers earned significantly greater tips while ovulating, but showed no similar increase if using oral contraceptives (Miller, Tybur & Jordan, 2007). All these indicators are passive.

Although field experiments involve active intervention by researchers in creating the conditions being studied, they often use outcome measures for which participants are unaware of being observed. Field experiments have been prominent in the bystander intervention literature, where the outcome is whether a helping intervention occurred. For example, in Piliavin, Rodin, and Piliavin’s (1969) classic experiment, a confederate feigning drunken behavior was less likely to receive help on a New York City subway train than a confederate feigning illness. Other studies have used the *lost letter technique*, in which fully addressed letters, varying according to the independent variable of interest (e.g., a return address of the Communist Party or the American Red Cross) are left in public places, to be found and mailed by passers-by, if they are so inclined (Milgram, Mann, & Harter, 1965). Other well-known field experiments in social psychology include manipulations of choice and responsibility in a sample of elderly nursing home residents, which, in an 18-month follow-up, were shown to have beneficially affected mortality rates (Langer & Rodin, 1976; Rodin & Langer, 1977). In still other studies, the number of available alternative choices influenced purchases of gourmet jams or chocolate (Iyengar & Lepper, 2000), drivers behind a stopped vehicle at a green light honked sooner if the stopped vehicle had a gun rack and an aggressive bumper sticker (Turner, Layton & Simons, 1975), men were no more likely to pay a return visit to a prostitute if she had played “hard to get” than if she had not (Walster, Walster, & Lambert, 1971), and, when French music was being played in a supermarket, French wine outsold German wine but when German music was being played, German wine outsold French wine (North & Hargreaves, 1997).

A Special Ethical Consideration in Field Experiments

Informed consent is a core principle of modern ethical regulations concerning the use of human participants in research. Even if participants in a laboratory study are not fully informed as a study commences, by their presence they have given consent, almost always explicitly, to participating in a study. This consent is based on an implicit and often explicit “contract” that expresses the participant’s willingness to be observed under experimentally created conditions, in return for the experimenter’s promise to protect his or her welfare and privacy. No such contract exists in field research. As described earlier, a prime rationale for field research is to examine natural behavior when people are unaware of being scrutinized. In many cases, asking potential participants in a field experiment to provide informed consent prior to a study would likely (and perhaps dramatically) reduce external validity.

Some commentators have argued for this reason that field experiments should be proscribed, but most Research Ethics committees allow some latitude. Regulations and their interpretation vary from one institution to another, although some generalizations are possible. Consent can typically be bypassed in studies that are solely observational and that involve anonymous, public behavior (e.g., pedestrian walking patterns). When interventions are involved and consent would interfere with external validity, researchers must take more than the usual amount of caution to ensure that participants will not be harmed, distressed, annoyed, or embarrassed. Practically, this means that field studies are typically limited to be less invasive than laboratory studies. (We suspect that few contemporary ethics committees would permit an experiment such as Piliavin et al.’s 1969 subway study, described earlier, because obtaining informed consent prior to the manipulation would render that study uninteresting.) Researchers can and should ask participants for consent and fully debrief them afterwards in most field experiments. Although post-hoc consent shows some degree of respect for participants’ privacy, it does not avert problems brought on by distress, embarrassment, or unwanted invasions of privacy. After-the-fact consenting may even alert participants that the situation just encountered was an experiment rather than a natural occurrence, potentially increasing negativity. Researchers and ethics committees therefore pay special attention to consent issues in field experiments.

Aronson et al. (1998) provide lengthier discussion of these issues.

INTERNET RESEARCH

In the late 1990s psychologists and other social scientists began using the Internet for research. At first the Internet

was simply a new medium for delivering conventional methods, most often surveys, to new populations in a cost-effective manner. For example, in 1996 one early study used an online form to collect pet owners’ ratings of their pets’ personalities (Gosling & Bonnenburg, 1998). Around the same time, Ulf-Dietrich Reips and John Krantz separately began using the Internet to deliver experiments to research participants (Musch & Reips, 2000). By today’s standards these early studies were rather rudimentary, and the samples were biased towards educated, technically savvy users. However, the studies hinted at the potential offered by the Internet. They showed, for example, that Internet studies could rapidly access large numbers of participants, many of whom were beyond the convenient reach of conventional methods, and they could do so at a fraction of the cost and without the laborious error-prone data entry associated with traditional methods. So if social psychologists were concerned about the critique of relying too heavily on convenience samples of college students (e.g., Sears, 1986), the Internet offered a ready solution.

It was not long before large-scale projects began to capitalize on the opportunities afforded by Web research, using Internet technology to improve the efficiency and accuracy with which traditional forms of data could be collected. In addition to reductions in data-entry errors, the Web allowed researchers to collect data around the world without the delays of land-based mail. Moreover, the validity of protocols could be checked instantly, the data stored automatically, and feedback delivered instantaneously to participants. This last benefit quickly proved to be particularly important because feedback served as a major incentive for participation (Reips, 2000). By providing personalized automated feedback, investigators were able to collect data from hundreds of thousands of participants, samples previously unheard of in psychological research. For example, since 1998 the Project Implicit website has collected several million tests of implicit attitudes, feelings, and cognitions from all over the world (<http://projectimplicit.net/generalinfo.php>).

The role of the Internet in psychological research has continued to expand as quickly as the growth of the Internet itself. An idea of the breadth of topics already covered by Internet research is conveyed by sampling the chapters of a volume summarizing recent trends in Internet psychology (Joinson, McKenna, Postmes, & Reips, 2007): In addition to well-studied areas of investigation, such as social identity theory, computer-mediated communication, and virtual communities, the volume also includes chapters on topics as diverse as deception and misrepresentation, online attitude change and persuasion, Internet addiction, online relationships, privacy and trust, health and leisure use of the Internet, and the psychology of interactive websites.

In recent years, the Internet has lived up to its promise of allowing researchers to access populations and phenomena that would be difficult to study using conventional methods. For example, to obtain access to white supremacists' attitudes about advocating violence toward Blacks, one group of researchers visited online chat rooms associated with supremacist groups (Glaser, Dixit, & Green, 2002). The researchers posed as neophytes, allowing them to conduct semi-structured interviews concerning the factors (threat type, threat level) most likely to elicit advocacy of violence. The anonymity afforded to both researchers and participants by the chat-room context and the easy access to a small, hard-to-reach group of individuals resulted in a dataset that would have been difficult to gather with conventional methods. Another study took advantage of the Internet to contact and survey a sample of people suffering from *sexsomnia*, a medical condition in which individuals engage in sexual activity during their sleep (Mangan & Reips, 2007); the embarrassment and shame experienced by sufferers meant that little was known about the condition. Yet, the reach and anonymity afforded by the Internet allowed the researchers to sample more than five times as many *sexsomnia* sufferers than had been reached in all previous studies combined from 20 years of research.

Domains of Web Research

In the first decade of the new millennium, Internet studies have proliferated, addressing a broad array of social psychological topics. To illustrate the scope of potential research strategies we next provide a non-exhaustive review of Internet-based studies.

The most basic class of Internet research—sometimes referred to as “translational methods” (Skitka & Sargis, 2006)—uses Internet technology to improve the effectiveness with which traditional forms of data can be collected. One prominent example of this approach is Project Implicit's large-scale administration of the Implicit Association Test (see Banaji & Heiphetz, this volume), which is designed to measure the strength of automatic associations between mental representations of various concepts (e.g., having implicit negative feelings toward the elderly compared to the young). And there have been many other successful attempts to measure attitudes, values, self-views, and any other entity that could formerly be measured with computers or paper-and-pencil instruments. Such studies are administered via computer, allowing them to take advantage of features associated with the medium, such as providing participants with immediate feedback, automatically checking for errors (e.g., missing responses), screening for invalid protocols (e.g., due to acquiescent responding), implementing adaptive testing (e.g., where the response to one stimulus determines

which stimulus is presented next), and presenting rich media (e.g., sounds and videos [Krantz, 2001; Krantz & Williams, in press]). Moreover, some methods that formerly involved cumbersome procedures, like sorting tasks, can be straightforwardly implemented online. For example, “drag-and-drop” objects can easily be used to complete ranking tasks, magnitude scaling, preference-point maps, and various other grouping or sorting tasks (Neubarth, in press; <http://hpolsurveys.com/enhance.htm>). As a result of these benefits, more and more researchers are doing basic experiments (e.g., on priming) via the Internet, sometimes delivering the studies no further than to a room on their own campus.

Many researchers have gone beyond merely using the Internet as a convenient and flexible way to deliver standard surveys, stimuli, and experiments to participants. Studies range from those that use the technological capabilities of computers connected to the Internet to gain access to new venues and populations (e.g., the White supremacists noted earlier) to those that focus on behavioral phenomena spawned by the Internet (e.g., Internet messaging, online social networking, large-scale music sharing).

The Internet provides opportunities to study phenomena unconstrained by the physical and practical parameters of the offline world. For example, personal websites can be used to examine identity claims that are hard to isolate in real-world contexts (Marcus, Machilek, & Schütz, 2006; Vazire & Gosling, 2004). Specifically, by exploiting the unique characteristics of personal websites and comparing personal websites with other contexts in which identity claims are made, the effects of deliberate self-expression can be isolated from the effects of inadvertent expression, which are confounded in most offline contexts of social perception. For example, a snowboard leaning against a bedroom wall may indeed reflect the occupant's past snowboarding behavior (i.e., behavioral residue), but her decision to leave it out rather than stow it in a closet could also reflect a deliberate statement directed to others about her lifestyle and preferences (i.e., an identity claim). In a physical room, one cannot tell whether the snowboard owes its presence to its role as behavioral residue, as an identity claim, or both. In contrast, most elements of a personal website have been placed there deliberately for others to see and the information on the sites can be rapidly saved and coded. (It is even possible to obtain records of past websites via www.archive.org, which is collecting them for the historical record). In a similar vein, the options of decorating and furnishing virtual spaces (e.g., in *Second Life*) are not subject to the practical, physical, and financial constraints associated with real-world spaces. The virtual world provides many more possibilities than those afforded by real life for experimenting with one's physical

representation (e.g., choosing avatars or game characters of a different sex, race, body type, and species).

In addition to being a domain in which to construct new studies to collect data, the Internet already contains rich pre-existing deposits of psychologically relevant data that vigilant researchers can harvest. For example, one study replicated findings derived from self-reported music preferences (which might be subject to self-reporting biases) with analyses of music libraries, which were accessible via a music-swapping website (Rentfrow & Gosling, 2003).

The millions of pages of text that are created online everyday provide another enormous source of pre-existing data. These pages offer opportunistic investigators an abundance of research possibilities. For example, as noted earlier, one project examining social psychological reactions to traumas analyzed the diaries of over a thousand U.S. users of an online journaling service spanning a period of four months, starting two months prior to the September 11th attacks (Cohn et al., 2004). Linguistic analyses of the journal entries revealed pronounced psychological changes in response to the attacks. In the short term, participants expressed more negative emotions, were more cognitively and socially engaged, and wrote with greater psychological distance. After two weeks, their moods and social referencing returned to baseline, and their use of cognitive-analytic words dropped below baseline. Over the next six weeks, social referencing decreased, and psychological distancing remained elevated relative to baseline. The effects were stronger for individuals highly preoccupied with September 11th but even participants who hardly wrote about the events showed comparable language changes. As noted by the authors this study bypassed many of the methodological obstacles of trauma research and provided a fine-grained analysis of the timeline of human coping with upheaval.

Another creative project used a German online auction site to examine ethnic discrimination (Shohat & Musch, 2003). The apparent ethnicity of sellers was manipulated by varying their last names. Analyses indicated that sellers with Turkish names took longer to receive winning bids than did those with German names. Given that so many interactions are now conducted online, and that many of them leave a trace, savvy researchers should be ready to pounce on opportunities as they arise.

An increasing number of studies focus on Internet behaviors as worthwhile social psychological phenomena in their own right, not simply because they are more convenient than studies done in the physical world. Some of these behaviors are extensions of offline behaviors but others are unique to the online world. With mobile Web access, Internet behaviors are becoming ever more integrated into the milieu of modern-day social interactions

and the distinction between online and offline life is becoming increasingly blurred; where, for example, is the line between speaking face-to-face, talking on the phone, and chatting via text or IM? With so much of contemporary social life played out online even those interactions that do not extend to offline contexts should be of interest to social psychologists because the laws of human behavior are likely to apply regardless of whether interactions are conducted on or offline.

By some estimates almost 600 million people worldwide have profiles on online social networking sites, such as MySpace and Facebook (<http://www.comscore.com/press/release.asp?press=2396>), making them an intriguing domain of inquiry. Which psychological needs are met by these sites? Which social psychological processes are operative? One early study of Facebook behavior examined how cues left by social partners on one's online networking profile can affect observers' impressions of the profile owner (Walther, van der Heide, Kim, Westerman, & Tong, 2008). The investigators examined the effects on profile owners of the attractiveness of people leaving "wall postings" (public notes left by friends on a person's profile page). Results suggested that the attractiveness of profile owners' friends affected ratings of their own attractiveness in an assimilative pattern, such that people with wall posts left by attractive friends were themselves viewed as more attractive than people with posts left by less attractive friends.

A large range of applications, such as online social networks, online role-playing games, and meeting software allow people to create online virtual representations of themselves (e.g., as game characters or avatars in virtual worlds). The advent of these representations creates whole new worlds for social psychological inquiry. For example, how are impressions formed and how are identities created in immersive virtual worlds such as those found in games like EverQuest, World of Warcraft, and in the virtual social network Second Life (Yee, Bailenson, Urbanek, Chang, & Merget, 2007)? And what are the connections between real people and their virtual representations? As more interactions and relationships become entirely virtual, it is important for researchers to examine the causes and consequences of the new social phenomena emerging in this domain.

The popularity of social networking sites and online multi-player videogames will almost certainly be superseded by new yet-to-be-invented online behaviors. Our point applies regardless: The online world is a legitimate venue in which to examine a plethora of social psychological behavior. Examples of online phenomena of potential interest to social psychologists include online message boards and chat rooms, Internet messaging (IM), virtual

worlds (e.g., Second Life), online support groups (e.g., for rare conditions), online multi-player video games (e.g., World of Warcraft), online social networks (e.g., Facebook), Internet dating (e.g., eHarmony), online auction sites (e.g., eBay), blogs, and an ever-growing list of others.

Overcoming Skepticism

Initial papers based on Internet research were greeted with a healthy dose of skepticism. Quite reasonably, journal editors and reviewers had a number of concerns about method artifacts and sampling issues. The major fears about Internet data can be summarized in terms of six concerns: (1) that Internet samples are not demographically diverse; (2) that Internet samples are maladjusted, socially isolated, or depressed; (3) that Internet data do not generalize across presentation formats; (4) that Internet participants are unmotivated; (5) that Internet data are compromised by the anonymity of the participants; and (6) that Internet-based findings differ from those obtained with other methods. These concerns were addressed in a study comparing a large Internet sample with a year's worth of conventional samples published in *JPSP* (Gosling et al., 2004). Analyses suggested that, compared to conventional samples, Internet samples are more diverse with respect to gender, socioeconomic status, geographic region, and age. Moreover, Internet findings generalize across presentation formats, are not adversely affected by non-serious or repeat responders, and are generally consistent with findings from traditional methods. Similar conclusions have been reached by other reviews addressing the validity of Internet research (e.g., Krantz & Dalal, 2000). As a result of these reviews and as Internet research has become more widespread, much of the skepticism has evaporated. Nonetheless, it is important to keep in mind the advantages and disadvantages associated with Internet-based methods.

Advantages and Disadvantages of Internet-Based Methods

As described earlier, Internet methods afford many advantages to social science researchers. The most important of these include the improved efficiency and accuracy with which traditional forms of data (e.g., surveys, informant reports, reaction-time experiments) can be collected, the possibility of instantly checking the validity of protocols and providing participants with immediate feedback, the ability to reach large and diverse samples from around the world, and the opportunity to integrate various media (e.g., sounds, photographs, videos) into studies (Gosling & Johnson, in press).

The central problems of Internet studies stem from the physical disconnect between researcher and participant, resulting in a potential lack of control over the assessment or experimental setting. Researchers are not physically present when Internet studies are conducted so they cannot easily assess participants' alertness and attentiveness. However, several methods have been developed to detect the degree to which participants are attending to the experimental materials and following instructions properly (Johnson, 2005; Oppenheimer, Meyvis, & Davidenko, 2008). For example, the Instructional Manipulation Check (IMC) measures whether participants are reading the instructions. The IMC works by embedding a question within the experimental materials that is similar to the other questions in length and response format but that asks participants to ignore the standard response format and instead provide confirmation that they have read the instruction (Oppenheimer et al., 2008).

Another potential problem with Internet studies is that researchers cannot easily answer questions from participants about the procedure. Because they are not directly observing research participants, researchers cannot be aware of possible distractions, such as eating, drinking, television, music, conversations with friends, and the perusal of other websites. Internet users, especially young Internet users, are notorious for multitasking while logged on, which could adversely affect the quality of Internet-based data. In the case of ability testing, with all of the information on the Internet at their disposal, it is difficult to keep participants from cheating. The extent to which these distractions and other available sources of information affect the findings of Internet studies is not known; however, research on Internet data versus real-life samples has allayed many concerns about data quality by showing that the Internet samples are generally not inferior to conventional samples from a psychometric standpoint (Gosling et al., 2004; Luce et al., 2007). Evidence is accumulating for their validity (Birnbau, 2004; Krantz & Dalal, 2000).

As noted earlier, one advantage of Internet research is its ability to reach samples beyond the reach of conventional methods. Internet-based samples tend to be more diverse and considerably more representative than the convenience samples of college students commonly used in psychology research (Birnbau, 2004; Gosling et al., 2004; Skitka & Sargis, 2006) but these samples are still not representative of the general population (Lebo, 2000; Lenhart, 2000). Participation in Internet-based research is restricted to people who have access to the Internet, know how to use a Web browser, and, in some kinds of research, have a functioning email address or instant messaging capability. People who are computer phobic, those who cannot afford a computer and Internet service and have no public access,

and those who are uninterested in learning how to browse the Web will be excluded from Internet research. Evidence is mixed regarding the extent to which this sampling bias affects the generalizability of Internet findings (Reips & Krantz, in press). Generally, and as in all research, investigators need to be cautious in making claims regarding the generalizability of their findings; to guide the scope of their generalizations, researchers should collect and report information about the demographics of their samples.

Finally, learning to construct Web pages, write program scripts, manage computer data bases, and engage in all of the other activities involved in starting up online research can be time consuming. Entire new sets of skills must be acquired, practiced, and polished. Fortunately, a large number of resources are available for aspiring Internet researchers, which we summarize next.

The Basics of Internet Research

The huge variety of possible topics, experimental designs, and implementation options make it impossible to provide much here in the way of specific advice on creating online experiments. Fortunately, a number of general books for investigators taking their first steps into the domain of Internet research are available (Birnbau, 2001; Fraley, 2004; Gosling & Johnson, in press), along with workshops (e.g., by Michael Birnbau or John E. Williams), and websites (e.g., iscience.eu; Project Implicit; websm.org). Birnbau (in press) provides a particularly useful introduction to the basic decisions that anyone planning to conduct an experiment online needs to make. These decisions range from deciding what kind of server makes most sense to choosing the appropriate client side (e.g., PHP, Perl) and server side (e.g., Java, JavaScript) programs and will be guided by design requirements. For example, JavaScript can be particularly useful for designs that require randomizing the order of materials or the assignment of participants to conditions, or adding checks for unreasonable or missing responses.

For researchers who do not want to program the websites themselves, several options exist, including survey websites (e.g., via Amazon's Mechanical Turk; Survey Monkey), websites for creating experiments (e.g., WEXTOR), collaborative opportunities with existing research groups (e.g., Project Implicit), and government-funded projects such as Time-sharing Experiments for the Social Sciences (TESS), which offers researchers opportunities to test their experimental ideas on large, diverse, randomly selected subject populations via the Internet.

Researchers who do not have the time or resources to program their own experiments from scratch will find WEXTOR ([http://psych-wextor.unizh.ch/wextor/en/index](http://psych-wextor.unizh.ch/wextor/en/index.php)

[php](http://psych-wextor.unizh.ch/wextor/en/index.php); Reips & Neuhaus, 2002) especially useful. WEXTOR is a free Web-based tool that allows researchers to quickly design and visualize a large variety of Web experiments in a guided step-by-step process. It dynamically creates the customized Web pages needed for an experimental procedure that will run on any platform and it delivers a print-ready display of the experimental design. Using an example of a 2 × 2 factorial design, Reips and Krantz (in press) provide a useful and accessible step-by-step description of how WEXTOR can be used to build a Web experiment; advice is also provided on how to monitor, manage, and reduce dropout rates (i.e., attrition). Ulf-Dietrich Reips also maintains [iScience.eu](http://iscience.eu), a free and up-to-date portal to many of the services useful for generating and editing experiments, recruiting participants, and archiving studies.

Many decisions face researchers undertaking studies on the Internet, and many potential pitfalls await the inexperienced or unwary investigator. Before undertaking Internet experiments, new researchers should draw on the numerous lessons already learned (e.g., Gosling & Johnson, in press; Reips, 2000, 2002a, 2002b, 2002c). A prudent first step would be to consult Reips (2002a), which summarizes expertise gleaned from the early years of Internet-based experimental research and presents recommendations on the ideal circumstances for conducting a study on the Internet, what precautions have to be undertaken in Web experimental design, which techniques have proven useful in Web experimenting, which frequent errors and misconceptions need to be avoided, and what should be reported. Reips's article concludes with a useful list of sixteen standards for Internet-based experimenting.

DIARY METHODS

Diary methods, also known as *event sampling*, have become increasingly popular and influential during the past three decades. A recent PsycInfo search revealed more than 1,200 published papers using or describing these methods. Although there is some flexibility in what counts as diary methods, they generally include measures for self-reporting behavior, affect, and cognition in everyday life, collected repeatedly over a number of days, either once daily (so-called *daily diaries*) or sampled several times during the day. The most popular of these latter sampling protocols are the *Experience Sampling Method* (ESM; Csikszentmihalyi, Larson, & Prescott, 1977) and *Ecological Momentary Assessment* (EMA; Stone & Shiffman, 1994; Shiffman, Stone, & Hufford, 2008). Another type of diary protocol is based on the occurrence of particular events, such as social interactions, sexual activity, or cigarette smoking.

Diary protocols are designed to “capture life as it is lived” (Bolger, Davis & Rafaeli, 2003, p. 580)—that is, to provide data about experience within its natural, spontaneous context (Reis, 1994). By documenting the “particulars of life,” researchers have a powerful tool for investigating social, psychological, and physiological processes within ordinary, everyday interaction. Key to the diary approach is an appreciation for “the importance of the contexts in which these processes unfold” (Bolger et al., 2003, p. 580) as a central element in the operation and impact of social psychological processes. As the accessibility and popularity of diary methods have grown, the kinds of questions that they can address have evolved in range and complexity.

Researchers have used diary methods to study a diverse range of phenomena and processes in social-personality psychology. Topics for which diary studies have become commonplace include affect (e.g., Conner & Barrett, 2005; Larsen, 1987; Sbarra & Emery, 2005), social interaction (e.g., Reis & Wheeler, 1991), marital and family interaction (e.g., Larson, Richards, & Perry-Jenkins, 1994; Story & Repetti, 2006), stress (e.g., Almeida, 2005), physical symptoms (e.g., Stone, Broderick, Porter, & Kaell, 1997), subjective well-being and mental health (e.g., Oishi, Schimmack, & Diener, 2001), and nearly every trait in the personality lexicon (e.g., Bolger & Zuckerman, 1995; Fleeson, 2004; Suls, Martin, & David, 1998). Other areas in which diary studies are less common but increasingly useful include sex (e.g., Birnbaum, Reis, & Mikulincer, 2006; Burleson, Trevathan, & Todd, 2007), self-esteem (e.g., Murray, Griffin, Rose, & Bellavia, 2003), self-regulation (e.g., Wood, Quinn, & Kashy, 2002), intergroup relations (e.g., Pemberton, Insko, & Schopler, 1996), social comparison processes (e.g., Wheeler & Miyake, 1992), social cognition (e.g., Skowronski, Betz, Thompson, & Shannon, 1991), attitudes (e.g., Conner, Perugini, O’Gorman, Ayres, & Prestwich, 2007), motivation (e.g., Patrick, Knee, Canevello, & Lonsbary, 2007; Woike, 1995), and culture and the self (e.g., Nezlek, Kafetsios, & Smith, 2008). For further details, we refer readers to surveys of diary methods used in social-psychological (Reis & Gable, 2000), psychopathology (deVries, 1992), and health psychology (Stone, Shiffman, Atienza, & Nebeling, 2007) research.

A Brief History of Diary Methods

Wheeler and Reis (1991) trace interest in the self-recording of everyday life events to four distinct historical trends in social science research: time-budget studies, which date back to the early 1900s (e.g., Bevens, 1913); the need in behaviorist therapies to have patients keep track of the behaviors being modified, such as smoking or marital conflict, so that treatment effectiveness could be

monitored (e.g., Nelson, 1977); industrial psychologists’ use of self-reports of work-related activity, as an adjunct to observation by outside observers (e.g., Burns, 1954); and checklist approaches to the study of life-event stress, popularized by Holmes and Rahe (1967). It was not until the seminal work of Csikszentmihalyi and colleagues (Csikszentmihalyi, Larson, & Prescott, 1977), who developed the ESM, that the field began to develop and apply systematic methods for studying everyday experience that could be adapted to diverse phenomena, questions, and circumstances. Csikszentmihalyi and his colleagues wanted to know more about the contexts in which *flow* (a mental state in which people are fully and energetically immersed in whatever they are doing) emerges, as well as its behavioral, affective, and cognitive correlates, and they felt that retrospective accounts were too inaccurate. Hence they decided to use pagers to randomly signal research participants, asking them to report on their experiences at the moment of the signal. At around the same time, Wheeler and Nezlek (1977) created the *Rochester Interaction Record* (RIR), a systematic method for recording the details of social interactions as they occur. More recently, Stone and Shiffman (1994) offered a similar method, EMA, which can incorporate physiological measures. Other sampling frameworks, notably including daily diary methods, in which respondents provide data once daily for a prescribed period of time, can be considered adaptations of these methods, although, as described below, the longer interval of a report and differing sampling schedule represents an important conceptual difference.

The Rationale for Diary Research

Diary studies have two main rationales, one conceptual and one methodological. The conceptual rationale is to capture information about daily life experiences, as they occur within the stream of ongoing, natural activity, and as they reflect the influence of context. Key is the idea that ordinary, spontaneous behavior, or what Reis and Wheeler called the “recurrent ‘little experiences’ of everyday life that fill most of our waking time and occupy the vast majority of our conscious attention” (1991, p. 340) can contribute to social-psychological knowledge. Two kinds of information fit under this heading. The first concerns basic facts: What happens when, where, and with whom else present. For example, diary studies can identify activity patterns, such as the relative distribution of studying, socializing, and TV watching among adolescents. The second refers to the *subjective phenomenology* of daily life: to examine “fluctuations in the stream of consciousness and the links between the external context and the contents of the mind” (Hektner, Schmidt, & Csikszentmihalyi,

2007, p. 6). Under this heading one might examine reports of affect and cognition, such as mood, focus of attention, self-evaluations, feelings of social connection, thoughts, worries, or wishes. Both kinds of information can be obtained with open-ended responses or with checklists and rating scales, although the latter is much more common in published research.

A more methodological rationale concerns the “dramatic reduction” (Bolger et al., 2003, p. 580) in the effects of retrospection, the result of minimizing the time between an event and its description. Traditional survey methods suffer from various well-researched biases, such as recency (more recent events are more likely to influence current judgments), salience (moments of peak intensity and distinctive or personally relevant events tend to be more influential), recall (the greater the time between an event and its recollection, the greater the potential distortion), state of mind (current states may influence recall of prior states), and aggregation (people find it difficult to summarize multiple events; see Reis & Gable, 2000; Hufford, 2007; Schwarz, Groves, & Schuman, 1998; Stone et al., 2000, for reviews). Diary methods are intended to reduce these biases as well as errors attributable to difficulty and to heuristic processing. This is a particularly central rationale for diary methods that require instantaneous reporting of what is going on at the moment that a signal is received (e.g., EMA, ESM). These biases are more likely to affect diary methods that cover longer periods (e.g., daily diaries or methods that ask for reports of events since the prior report), although the extent of such effects, which depends on the time gap, the questions being asked, and the nature of the events, is likely to be less than with traditional surveys.

Diary methods also have certain advantages over observational methods that sample a narrower range of behavior, such as laboratory observations of dyadic interaction. Although laboratory observations provide a videotaped record that, with considerable time and effort, can be coded from an independent and semi-objective perspective, the structured context of being observed by experts in a restrictive setting may elicit behavior that is unrepresentative of more natural, unstructured settings (Reis, 1994). (For example, participants in a laboratory observation typically cannot get up and turn on the TV, as they can during real-life conflicts.) Furthermore, observational studies rarely provide information about behavior in more than one or two contexts, whereas diary studies can be informative about multiple and diverse contexts, a key consideration for studies seeking to identify contextual determinants of behavior.

Types of Questions For Which Diary Methods Are Well Suited

Perhaps understandably, given their history, diary methods have had appeal for descriptive research. For example,

diary methods have documented how people spend their time (Robinson & Godbey, 1997), with whom they socialize (Reis & Wheeler, 1991), what they eat (Glanz, Murphy, Moylan, Evensen, & Curb, 2006), and how often they feel bored (Csikszentmihalyi et al., 1977). Behavior description is an important, under-recognized and under-practiced component of theory development in social-personality psychology (Rozin, 2001, 2009). More than a half-century ago, Solomon Asch argued, “Before we inquire into origins and functional relations, it is necessary to know the thing we are trying to explain” (1952, p. 65). McClelland (1957) made a similar argument for personality theory, suggesting that behavioral frequencies may be the best place for personality theorizing to begin. As Reis and Gable commented, “to carve nature at its joints, one must first locate those joints” (2000, p. 192).

Nonetheless, social-personality psychologists are most likely to apply diary methods for testing theory-driven hypotheses in three different ways. First, diary methods can be used to evaluate alternative mechanisms thought to underlie an effect. For example, comparing three potential explanations for the observed correlation between trait neuroticism and distress, Bolger and Schilling (1991) reported the best support for the tendency of persons high in neuroticism to react more strongly to stressful circumstances. Second, diary studies can be used to distinguish competing predictions. An example is Wheeler and Miyake’s (1992) contrast of two plausible effects of mood on subsequent social comparison: cognitive priming, which predicts comparing upward, to better-off persons, and self-enhancement, which predicts downward comparison, to less-fortunate others. Upward comparison was better supported. Third, diary methods are particularly well suited to identifying conditions under which effects vary in strength or relevance (moderators). For example, solitary drinking is more likely on days with negative interpersonal experiences, whereas social drinking is more likely on days with positive interpersonal experiences (Mohr et al., 2001).

Looking at the types of questions that diary methods can address in a somewhat different way, Bolger et al. (2003) described three types of research questions to which diary studies are suited: *aggregating over time*, *modeling the time course*, and *examining within-person processes*. The first asks about persons over time and context and involves aggregating individual responses over multiple reports. Typically, this approach is meant to improve over methods that ask respondents to summarize their experience during a timespan (e.g., “How much have you socialized with others during the past two weeks?”), which are subject to retrospection, selection, and aggregation biases. Although diary designs sometimes seem like overkill for questions of this sort—asking people to repeatedly report a range of

experiences for the purpose of arriving at a single summary score—the substantial increase in data quality provides more than adequate justification for the effort. The opportunity for “data mining”—sorting through large amounts of data to ask more refined, more detailed, or alternative questions—is an additional tangible benefit.

Modeling the time course allows researchers to explore temporal and/or cyclical patterns in phenomena. Well-known among these patterns are diurnal (Clark, Watson, & Leeka, 1989) and weekly (Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Stone, Hedges, Neale, & Satin, 1985) cycles of affect, such that positive affect tends to be higher, and negative affect lower, in the early evenings and on weekends, respectively. Diary designs are also amenable to identifying more complex trends (e.g., repeated “up and down” cycles, such as might be shown in a sine wave [Walls & Schafer, 2006]), longer intervals (e.g., seasons or years), dynamic models, or so-called “broken stick” or step-function models, in which the pattern of an outcome variable is discontinuous before and after a particular point (e.g., following a major life event, such as September 11th, unemployment, or divorce). Analyses of this sort have been rare in social psychology.

The most widespread use of diary designs in social psychology falls into Bolger et al.’s (2003) third category, examining within-person processes. Such studies investigate “the antecedents, correlates, and consequences of daily experiences” (Bolger et al., 2003, p. 586) as well as, potentially, the processes underlying their operation. For example, studies have shown that high work stress is likely to lead to family conflict (Repetti, 1989) and that invisible support tends to yield better adjustment to stressors than visible support does (Bolger, Zuckerman, & Kessler, 2000). Many researchers construe this use of diary methods as the non-experimental equivalent of experimentation, inasmuch as the association between specified independent and dependent variables can be assessed. However, there is an obvious and important difference: Experiments involving random assignment of participants to conditions permit causal inference, whereas diary studies do not (although data analyses can rule out some alternative explanations, as described below). Conversely, a major advantage of diary methods is their ability to examine Person \times Environment ($P \times E$) interactions, or whether situational effects vary systematically for different kinds of persons. For example, low self-esteem persons respond to perceived relationship threats by distancing from their partners whereas high self-esteem persons respond to the same kind of threats by moving closer (Murray et al., 2003). By allowing researchers to track individual differences in response to variability in the natural environment, diary

methods are ideal for studying $P \times E$ effects of the sort first theorized by Lewin and since then endorsed, at least in the abstract, by nearly all social and personality psychologists (Fleeson, 2004; Funder, 2006).

Diary designs also have the important advantage of unconfounding between-person and within-person questions. Consider the hypothesis that perceived discrimination is associated with lower effort in achievement settings. This might be studied by characterizing a person’s experiences with discrimination (e.g., with a questionnaire) and relating those scores to measures of achievement-related effort. An alternative study might sample moments in a person’s life, assessing ongoing covariation between perceived discrimination and achievement-related effort. Although seemingly similar, these two hypothetical studies address independent questions. The former study asks a personological question: Do persons who tend to perceive discrimination also exert differential effort in achievement-related settings? The latter study asks an experiential question: When discrimination occurs, do people respond with differential effort? Numerous theorists (e.g., Epstein, 1983; Gable & Reis, 1999) have noted that these questions, and hence the nature of the processes that would explain their answers, differ fundamentally. In a more general way, Campbell and Molenaar (in press) argue that much of psychological science erroneously assumes that intra-individual variation in response to time or context follows the same rules and mechanisms as inter-individual variation. They discuss what they see as a major reorientation in the field toward “person-specific paradigms,” capable of distinguishing these different levels of explanation. Diary methods are a powerful tool for any such reorientation.

Design and Methodological Issues in Diary Research

Like any research paradigm, diary methods require that researchers make choices guided by conceptual and practical concerns. Diary methods are flexible and can be tailored to the needs of an investigation. At the same time, planning and conducting research requires addressing inherent practical issues and limitations. Below we review some of the more important (and in some cases contentious) issues that have arisen in current practice. More detailed information is available in Christensen, Barrett, Bliss-Moreau, Lebo, and Kaschub (2003), Conner, Barrett, Tugade, and Tennen (2007), Reis and Gable (2000), or Christensen’s website, <http://psychiatry.uchc.edu/faculty/files/conner/ESM.htm>. Hektner et al. (2007) describe practical issues in the ESM in more detail, and Piasecki, Hufford, Solhan, and Trull (2007) describe the application of diary methods in clinical assessment. For a proposed list of methodological

information to include in journal reports, see Stone and Shiffman (2007).

Designs

Choice among reporting protocols is generally based on two considerations: The research question and the relative frequency of the key phenomena. Wheeler and Reis (1991) described three major protocols: *interval-*, *signal-*, and *event-contingent* (see also Reis & Gable, 2000). Interval and signal are preferred choices when researchers are interested in “phenomena as they unfold over time” (Bolger et al., 2003, p. 588) or when the phenomena occur often. Studies that focus on specific events, especially rare events (e.g., major life events, drug use), are more likely to use event-contingent protocols.

Interval-contingent methods require reports at regular, predetermined times, so that the gap between successive reports is relatively constant. The most popular such schedule is the daily diary, in which participants report on their experiences once a day, typically in the evening, before bedtime. This daily cycle is consistent with the importance of the day as a natural interval for organizing life activities, as well as circadian rhythms underlying certain biological and psychological processes (e.g., DeYoung, Hasher, Djikic, Criger, & Peterson, 2007; Hasler, Mehl, Bootzin, & Vazire, 2008). Other research has collected reports at several fixed times during the day, such as thrice-daily (noon, dinnertime, and bedtime; Larsen & Kasimatis, 1991). Fixed intervals are particularly valuable for studies of time-sequences and temporal cycles, in which repetitive, constant, or precisely timed intervals are helpful or essential (e.g., day-of-the-week effects or the time-bound impact of activities, such as meals or afterschool activities).

Signal-contingent protocols prompt participants to report their experiences at the moment of receiving a signal, usually delivered by pagers, cell phones, or preprogrammed devices (e.g., palmtop computers, watches). As a rule, signaling schedules are random and unpredictable within predetermined blocks of time, so that a fixed number of prompts are sent each day (often, but not necessarily, around 10). Randomness is key: Because the data presumably represent a random sampling of daily experiences, researchers receive non-selective, unbiased estimates of the distribution and quality of daily activities, affects, and cognitions. Non-random signals might be skewed toward particular kinds of experiences, and predictability would allow participants to alter their activities shortly before signal. Signal-contingent methods also typically demand that participants report their experiences right at the moment of signal, with little or no delay, to support the claim that they represent “real-time data capture” (Stone et al., 2007) without functional retrospection bias.

Signal-contingent protocols are limited in their ability to capture rare, occasional, or fleeting events, for which *event-contingent* protocols are better suited. When events are rare or short-lived, random signals are unlikely to sample a sufficient number, especially when researchers wish to compare different subtypes of those events. Thus, instructions may ask participants to record their experiences whenever a target event occurs. Examples include social interactions lasting 10 minutes or longer (Wheeler & Reis, 1991), conflicts among adolescents (Jensen-Campbell & Graziano, 2000), ostracism (Williams, 2001), sex (Birnbaum et al., 2006), smoking (Moghaddam & Ferguson, 2007), alcohol consumption (Mohr et al., 2001), altruistic thoughts and behavior (Ferguson & Chandler, 2005), social comparisons (Wheeler & Miyake, 1992), prejudice and discrimination (Swim, Hyers, Cohen, & Ferguson, 2001), and stressful events (Buunk & Peeters, 1994). To avoid bias, event-contingent protocols must unambiguously define the types of events to be reported, and participants must do so when those events occur. Alternatively, a recent technological innovation uses sensors embedded within a recording device to detect certain events (for example, an accelerometer to monitor activity levels) and signal the respondent to provide a report (Choudhury et al., 2008).

Delivery Systems

The earliest ESM studies, reflecting that era’s technology, used pagers to prompt participants to complete paper-and-pencil records. Since then, advances in microprocessing technology have enabled many more sophisticated systems for collecting diary data. One of the earliest developments relied on digital watches, which could be preprogrammed to deliver on schedule a week or more’s worth of signals, although paper-and-pencil reports were still required (Delespaul, 1992). A more important advance came from Personal Digital Assistants (PDAs, such as palmtop computers), which allowed researchers to signal participants, collect responses, and branch to different question sets depending on what the participant is doing at the time (Barrett & Barrett, 2001). Several websites provide or describe programming for such devices, at least one of which, developed by Lisa Feldman Barrett and Daniel Barrett with National Science Foundation support, is free (www2.bc.edu/~barretli/esp). Relative to paper-and-pencil, PDAs offer the advantage of verifying the time of the participant’s response (which can then be compared to the schedule to assess fidelity, as described below) and can also record the reaction time or duration of responding for particular questions. On the other hand, PDAs are costly, breakable, stealable, and can be difficult, inconvenient, or intrusive for some participants (e.g., with elderly or less tech-savvy samples) and in

some circumstances (e.g., when participants are in classes or meetings). Some researchers have created specialized or proprietary devices that can be programmed to accommodate particular circumstances (e.g., Invivodata Inc). A relatively recent and promising development uses cell phones in this manner (Collins, Kashdan, & Gollnisch, 2003). Downloadable software for using cell phones to conduct context-aware experience sampling can be found at <http://myexperience.sourceforge.net>.

The regularity of interval-contingent protocols permits use of dedicated Internet sites for data collection. These tend to be appropriate when participants have easy Internet access (e.g., college students), and the scheduled timing of reports is consistent with this access (e.g., end of the day, when participants are at home). Internet data collection verifies the time of reporting and has the further advantage of allowing researchers to monitor compliance in real time, so that noncomplying participants can be contacted immediately. A conceptually similar low-tech approach involves a telephone call to participants at each prearranged time and having an interviewer ask questions and record answers (Wethington & Almeida, in press). Verification of the time of report is particularly important when researchers wish to synchronize diary reports with other ambulatory measures, such as physiological data. For example, one study examined covariation in cardiac function and emotional experience at randomly selected moments over 3 days (Lane, Reis, Peterson, Zareba, & Moss, 2008).

The benefits of electronic data collection methods notwithstanding, many researchers (including us) still see an appropriate role for paper-and-pencil diaries. When electronic methods are not needed to deliver random signals or varying protocols, and when the need to verify compliance is either not great or otherwise achievable (see below), paper-and-pencil diaries (e.g., in booklet form) are convenient, easy, accessible, user-friendly, and minimally burdensome, all significant advantages when people are asked to report repeatedly in a personal way on their activities, thoughts, and feelings. We therefore recommend that the choice of delivery systems take into account both the needs of the research and the likely experience of participants when using that system.

Fidelity

Because the rationale for diary studies depends on timely reporting, controversy exists about whether participants can be trusted to comply with these schedules in the absence of verification. This controversy was made prominent in an important study comparing compliance among participants using an electronic diary, which overtly recorded the time of response, with paper diaries contained in a logbook that surreptitiously recorded openings and

closings³ (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). About 94% of the electronic diaries were compliant with the reporting schedule, but only 11% of the paper responses. A further problem was that the vast majority of the paper-condition participants claimed (apparently falsely) that they had been compliant. Chief among the factors that may contribute to noncompliance is *hoarding*: the tendency to complete multiple records at one time, such as shortly before collection by researchers.

Most published research either cannot or does not verify compliance, although diary researchers would likely agree that noncompliance varies across studies, contexts, and persons. Nevertheless, subsequent studies have suggested that the problem of noncompliance may not be as pandemic as Stone et al. indicate. For example, three studies reported by Gaertner, Elsner, Pollmann-Dahmen, Radbruch, and Sabatowski (2004) indicate that noncompliance is far less common than Stone et al. report, a conclusion similar to that of Tennen, Affleck, Coyne, Larsen, and DeLongis, who state, "in six separate studies, we found almost no evidence of hoarding" (2006, p. 115). To social psychologists, a more important question than the frequency of noncompliance is the question of impact. In this regard, Green, Rafaeli, Bolger, Shrout, and Reis (2006) conducted extensive analyses, concluding that paper diaries (which could not be verified) and electronic diaries (which could be verified) yielded psychometrically equivalent data and findings. A similar conclusion follows from another study comparing electronic and paper pain diaries (Gaertner et al., 2004).

Perhaps more than with most methods, we see fidelity as a matter of participant motivation: Diaries are often burdensome, and they require that participants regularly and reliably invest a significant amount of time and attention to describing their experiences. Client motivation affects patient compliance with self-reporting protocols in behavior therapy research (Korotitsch & Nelson-Gray, 1999). For this reason, many diary researchers emphasize the importance of developing a collaborative, trusting relationship with participants. It is unlikely that the fact of monitoring or the method of administration—e.g., using a PDA that records time of response—will resolve most issues of noncompliance. For example, if participants are busy, have misplaced their PDA, become reactive to the suggestion that they cannot be trusted, feel that the protocol is difficult or unpleasant, or do not feel enough commitment to the research project to prioritize timely recording, noncompliance rates may be high. (Simply eliminating

³ An unfortunate confound in this study is that there were also other differences between conditions.

persons or reports exceeding some compliance criterion may introduce nonrandomness, a potentially important problem.) Furthermore, even near-immediate reports are not free of memory-related distortion (Takarangi, Garry, & Loftus, 2006). For these reasons, researchers should take steps to minimize the motivation and opportunity for noncompliance rather than emphasizing monitoring. When objective verification is desired, PDAs or Internet sites routinely record time of response. Compliance can be monitored with paper diaries, such as with a portable secure (unalterable) time-stamping device or, for daily diaries, by requiring that data be handed in or mailed each day. Postmarks might also be used as an admittedly imperfect variant of the *bogus pipeline* (a technique for reducing response bias whereby research participants are led to believe that researchers have access to their true feelings or attitudes) for encouraging and monitoring compliance (Tennen et al., 2006).

Reactivity

Researchers sometimes worry that the process of diary record-keeping may alter participants' experiences and reports. Hypothetically, any of several effects are feasible. Self-monitoring might enhance awareness of personal behavior—for example, eating or work habits—motivating participants to pursue change. Self-awareness may reduce the intensity of affective states (Silvia, 2002) and introspection about traumatic events may facilitate healthy cognitive reorganization (Pennebaker, 1997). Habituation or response decay over time might lead to stereotyped, non-thoughtful responding. Knowledge about a phenomenon—for example, which circumstances seem to be associated with memory loss—might develop as participants reflect on their personal experiences with it. Anticipation of a diary report might even cause participants to modify their behavior. For example, asking people about their intent to engage in certain behaviors increased the frequency of those behaviors in three nondiary studies (Levav & Fitzsimons, 2006). Similarly, participants might avoid undesirable or illegal activities, or circumstances that will be effortful to describe, lest they have to inform researchers about those activities.

Although little research has investigated these possibilities, what research there is suggests minimal problems. Some studies report little effect of repeated responding (e.g., Hufford, Shields, Shiffman, Paty, & Balabanis, 2002), whereas other studies have found small effects as a function of the number of required reports (Mahoney, Moura & Wade, 1973) or the obtrusiveness of the recording process (Kirby, Fowler & Wade, 1991). The process of recording healthy habits had no discernible effects on enactment of those habits (Conti, 2000), nor did keeping diaries of

marital conflicts for 15 days alter spouses' behavior on a videotaped conflict-resolution task (Merrilees, Goeke-Morey, & Cummings, 2008). Similarly, momentary reports of mood collected several times a day did not enhance later recollection of those moods (Thomas & Diener, 1990). And, although a sample of treated alcoholics claimed becoming more aware of their drinking patterns after taking part in a signal-contingent protocol, few actual differences were observed (Litt, Cooney, & Morse, 1998).

These reassuring findings notwithstanding, the potential for reactivity problems suggests the need for caution in designing protocols, minimizing factors that may adversely affect participants' willingness to be thoughtful and specific (e.g., asking too many similar questions; insensitivity to interference with normal activities; running studies for unnecessarily lengthy periods). Analyses should also routinely examine data for signs of response stereotypy or carelessness, or for changes in the nature and pattern of responses from early and late records (e.g., comparing week-1 and week-2 means and variances in a two-week diary study; see Green et al., 2006, for examples). Finally, we concur with others (e.g., Bolger et al., 2003; Gable & Reis, 2000; Rafaeli, 2009) who have called for further research into reactivity effects. Such research would have methodological benefits, and would shed light on the role of self-monitoring and awareness in everyday experience.

Data Analytic Considerations

Diary data represent an analytic challenge for two reasons. Statistically, diary data are nested within individuals (and often individuals are nested within higher-order units, such as couples, classrooms, or work groups), so that repeated observations are not independent (Kenny, Kashy, & Bolger, 1998). Furthermore, as with most time-series data, variables in one report are likely to be correlated with prior reports, creating *autocorrelation*, which must be considered in data analyses. Conceptually, because researchers using diary methods are usually interested in something more than a count of the total number of stressful events or the mean level of intimacy across all social interactions, at the least simple aggregation underutilizes effortfully collected, potentially informative data.

Multilevel models have become the standard method of analysis, allowing researchers to examine both between-person and within-person processes (that is, variation within person as a function of time or conditions), as well as their interaction. Although under certain circumstances multilevel analyses can be conducted with traditional methods (e.g., repeated-measures analysis of variance), maximum-likelihood estimation is more common. Maximum-likelihood methods (e.g., Hierarchical Linear Modeling [Bryk & Raudenbush, 1992]) provide more

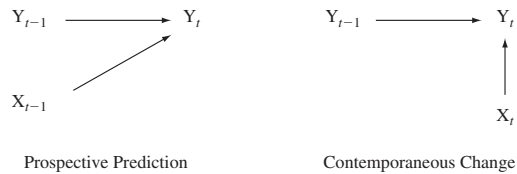


Figure 3.2 Two Kinds of Temporal Comparisons in Diary Designs.

accurate estimation of population values, especially when the number of records varies from one person to the next and when random effects are considered more appropriate than the usual fixed effects. Excellent discussions of these analytic methods are available elsewhere (e.g., Bolger et al., 2003; Nezlek, 2003; Schwartz & Stone, 1998; Walls & Schafer, 2006; West & Hepworth, 1991), so we do not discuss them here.

Given the possibility of carryover from one report to the next, researchers often analyze a given criterion variable by controlling for the prior report's value of that variable—for example, by examining today's affect controlling for yesterday's affect. This is commonly done in either of two ways, as shown in Figure 3.2, and their implications differ significantly, although the choice is rarely explicit. The first method, prospective prediction, involves analyzing the outcome variable on a given day t as a function of the predictor and outcome on the prior day, $t-1$. The second method, contemporaneous change, looks at covariation between outcome and predictor on a given day t controlling for the outcome on the prior day $t-1$. The major rationale for prospective prediction concerns inferences about causal priority. By predicting outcomes from both prior-day variables, reverse causality—that the outcome is causally responsible for the predictor—is rendered implausible. In other words, and similar to the logic of prospective prediction in longitudinal studies, because the partialled predictor at time $t-1$ shares no common variance with the outcome yet temporally precedes the outcome at time t , it plausibly exerts a causal effect on the outcome. For example, this method has been used to establish that daily events are more likely to be causally responsible for daily affect than the reverse (Gable, Reis, & Elliot, 2000). On the other hand, because in the contemporaneous change model outcome and predictor are assessed simultaneously, causal priority cannot be ascertained. However, controlling for the prior $t-1$ outcome variable removes carryover effects so that whatever associations are obtained result from that moment or interval, rather than prior moments or intervals.

Although prospective prediction has clear advantages, there is a potential downside: The effects of the predictor

variable must be durable enough to persist from reports at $t-1$ to reports at t . This seems more likely in designs where assessments are separated by relatively small intervals (e.g., ESM, EMA). In the common daily diary designs, prospective prediction requires that effects endure from one day to the next, a relatively tenuous assumption for many phenomena, given that a full day's worth of activity, as well as the restorative effects of sleep, intervene. It follows furthermore that in the absence of intervening events, the contemporaneous change model provides a more accurate estimate of the association between outcome and predictor. For this reason, contemporary change models are preferable in certain instances, their greater inferential ambiguity notwithstanding. The choice of analytic models, therefore, should be based on the researcher's goals.

Although social psychologists have been quick to adopt diary methods for examining processes within persons, they have been slow to use these methods for investigating more complex temporal patterns. For example, one might use spectral analysis to examine the periodicity (frequency and amplitude of repetitive cycles) of various phenomena, such as mood, over the day (Larsen, 1987) or week (e.g., the day-of-the-week effect; Reis et al., 2000), or in response to major life events (e.g., bereavement). Investigating the natural life cycle of phenomena such as conflict, instances of ostracism or discrimination, affective forecasts, or persuasive appeals, and accounting for variability in these cycles as a function of situational factors and individual differences is a fertile opportunity for expanding social psychological knowledge. Another type of analysis exploits the repeated sampling of diary designs by using temporal models to specify processes that contribute to continuity and discontinuity in social behavior over time. In this regard, Fraley and Roberts (2005) propose different statistical models that contribute to longitudinal stability—that is, to a high test-retest correlation—in personality characteristics over the life course. These models can also be used to better understand stability in social-psychological phenomena over shorter intervals.

Diary Research with Couples and Families

Diary methods, especially daily diaries, have become particularly popular among researchers who study couples and families. All of the advantages of diary methods discussed earlier apply to couples and families; additionally, diary methods allow researchers to study interactive processes (e.g., family conflict, intimacy) as they unfold in interdependent social units and also to identify contextual and dispositional factors that moderate their impact (e.g., work stress, self-esteem). For example, one partner's feelings of vulnerability may engender behavior that contributes to the other partner's dissatisfaction with the relationship, a

process that is exacerbated when the vulnerable partner is high in rejection sensitivity or low in self-esteem (Downey, Freitas, Michaelis, & Khouri, 1998; Murray et al., 2003).

Conducting diary research with couples and families generally necessitates that partners do not discuss their responses and that they keep their reports confidential from each other. Confidentiality is important because partners might well be reluctant to report certain behaviors (e.g., violence, infidelity, sources of dissatisfaction) if there was even a slim chance that their partners might see their reports. Privacy can be difficult to ensure with standard delivery systems, so that dedicated systems are preferred (e.g., cell phones or PDAs that do not store responses locally or that are password-protected). The former is particularly important when comparisons of partners' perspectives are of interest, as in the example of studies that examine the relative impact on daily affect and relationship well-being of shared and differing perspectives about everyday couple interaction (Gable, Reis, & Downey, 2003). At the same time, couple and family researchers coordinate reporting schedules so that all parties provide reports at the same time or following the same events. Otherwise, one would not know if divergence reflected differing perspectives on the same interaction or whether different interactions were being described.

Couple and family data require special methods of analysis to manage interdependence (Kenny, Kashy, & Cook, 2006), and multilevel analyses of diary data are no exception. The couple/family adds an additional level of nesting to such analyses (repeated reports are nested within individuals, whose responses are nested within the couple or family), and some researchers prefer to analyze these data as three-level hierarchical models. Nevertheless, there are both practical and statistical reasons to consider using two-level models, using a technique introduced by Raudenbush, Brennan, and Barnett (1995) and recently described by Laurenceau and Bolger (2005). This method takes advantage of the fact that partners are distinguishable—for example, one is husband and one is wife—so that predictor variables representing both of them can be included in the same level of analysis.

AMBULATORY ASSESSMENT

The term *ambulatory assessment* refers to the use of mechanical or electronic devices to record information about an individual's activity, circumstances, or states within ordinary daily life.⁴ First developed for medical purposes—specifically, monitoring of blood pressure and cardiac function over the course of normal activity—with the increasing complexity and miniaturization of digital

technology, ambulatory methods have become exponentially more useful and adaptable to research. In this section we briefly review the application of these methods to social psychology. Fahrenberg and Myrtek (2001) provide a more general review. It bears noting that most researchers include momentary self-report methods, such as ESM and EMA, under the general heading of ambulatory assessment. This chapter has not followed that convention because ESM and EMA are used primarily for self-reports, whereas the methods reviewed in this section are non-self-report. Non-self-report ambulatory methods can also be combined with ESM and EMA, as several examples below show.

The main rationale for ambulatory assessment is the same as that discussed earlier for diary methods: To provide detailed data about variables of interest within their natural, spontaneous context. By applying this approach to behavioral (i.e., not self-reported) data, researchers capitalize on the advantages of non-laboratory assessment—external validity, repeated, contextually sensitive data—while avoiding the pitfalls of self-reports (Stone et al., 2000). Ambulatory measures are particularly useful for assessing processes that operate outside of awareness, which cannot be self-reported. Currently available ambulatory methods include tools for assessing physiological processes, location and activity, speech, and features of the ambient environment. How this is done varies markedly. Ambulatory measures can be obtrusive (e.g., blood pressure monitors) or unobtrusive (e.g., sound recording devices, motion sensors), and they can be self-contained (e.g., PDAs) or telemetric (devices that transmit data remotely, such as by using mobile phone technology). As modern technology has expanded the range of what is possible, more and more sophisticated gadgets and gizmos have been designed with significant potential for behavioral science research (Goodwin, Velicer, & Intille, 2008). Application of these tools in theory-oriented research programs has been variable, with some gaining immediate favor and others awaiting adoption. This variability reflects several practical considerations—cost, ease of use, adaptability to particular circumstances, involvement of social psychologists in development and dissemination—as well as a more

⁴ Many researchers include ESM and EMA in the general category of ambulatory assessment methods, because many of the same methodological and conceptual principles mentioned here also apply to ESM and EMA. We do not follow that convention for two reasons. First, common practice in social psychology uses ESM and EMA in much the same manner as other diary methods. Second, the data collected with ESM and EMA are self-reports of experiences, thoughts, and feelings, much like diary data, whereas we limit the discussion of ambulatory assessment methods to direct recording of non-self-report data.

fundamental question: Researchers need to imagine how a new method can enhance the informativeness of their work. In some instances, technological advances offer relatively small potential for theoretical advances, whereas in other instances, these advances may have potential to dramatically improve the quality and relevance of findings. In still other instances, a new device may open an entirely new area to social-psychological research.

Below we describe four examples—two established, two novel—with particular relevance to social psychology.

Ambulatory Cardiovascular Monitoring

Ambulatory blood pressure monitoring for medical purposes is now commonplace, as research showed that blood pressure recordings taken in the individual's normal environment were better indicators of cardiovascular risk than office-based assessments (e.g., White, Schulman, McCabe, Holley, & Dey, 1989). Cardiovascular reactivity has long been considered an important marker of stress, and more particularly of whether stressful circumstances are appraised as threatening or challenging (Blascovich, 2000). Combining these two principles suggests that ambulatory cardiovascular monitors would provide better indications of the impact of social-personality factors on cardiovascular function than laboratory assessments. For example, trait negative affects (depression, anger, neuroticism) predict higher blood pressure in daily life (Ewart & Kolodner, 1994; Raikkonen, Matthews, Flory, Owens, & Gump, 1999). In a related vein, lonely people tend to be higher than non-lonely persons in total peripheral resistance, a physiological indication of threat responses (Hawley et al., 2003).

If ambulatory measurements are combined with event records (such as daily diaries), within-person changes to social-psychologically relevant events can be assessed. Thus research has shown that the association between trait negative affectivity and blood pressure elevation is stronger in the classroom than in other settings (Ewart & Kolodner, 1994), and that New York City traffic enforcement officers experience higher blood pressure than baseline when engaging in unpleasant communications with the public (Brondolo, Karlin, Alexander, Bobrow, & Schwartz, 1999). Similarly, momentary negative moods elevated both systolic and diastolic blood pressure among optimists to approximately the same levels characteristic of chronically negative people (Raikkonen et al., 1999).

Ambulatory cardiovascular monitors have become increasingly sophisticated and are now capable of recording more than blood pressure and vascular resistance. For example, Holter monitors can continuously record cardiac activity (much like an office ECG) for periods as long as 24

hours. Using a sample of individuals with Long QT syndrome (a genetically based disorder that puts affected individuals at risk for sudden cardiac death), Lane et al. (2008) found that positive emotion was associated with shortening of the Q-T interval, lessening the risk of cardiac events.

Electronic Recording of the Acoustic Environment

Observational researchers sometimes fantasize about implanting recording devices on the person of a research participant so as to obtain an objective account of everything they do. The *Electronically Activated Recorder* (EAR), developed by Pennebaker and colleagues (Pennebaker, Mehl, Crow, Dabbs, & Price, 2001), represents a first, less megalomaniacal step in that direction. The EAR is a portable audio recorder that unobtrusively switches on and off at random or regular intervals, providing samples of the acoustic environment as participants go about their daily activities. Participants cannot tell when the device is recording, allowing researchers an opportunity to unobtrusively observe even relatively subtle sounds. For example, in several studies, the EAR switched on for 30 seconds every 12.5 minutes, yielding about 70 samples per person per day, which contain about 35 minutes worth of recordings (Mehl, Vazire, Ramirez-Espinosa, Slatcher, & Pennebaker, 2007). Most commonly, researchers have used the EAR to sample spoken language (i.e., verbal content and linguistic styles, which can be transcribed and analyzed via manual content analysis or text-analysis software), but it can also be used to describe the acoustic environment in other ways; for example, what sorts of interactions or activities are taking place (Mehl & Pennebaker, 2003a). An added benefit is that EAR transcriptions are easily archived for subsequent reanalysis as new hypotheses emerge.

The EAR has been used in social psychology in several ways. One analysis reported simple word frequencies from six studies, concluding that the popular stereotype that women are more talkative than men is unfounded (Mehl et al., 2007). Men and women both used about 16,000 words per day—with large individual differences but no evidence of a sex difference (the least talkative person used 695 words, the most talkative 47,016 words). A study fortuitously begun just before the events of September 11, 2001, found that a relative preponderance of dyadic over group interactions fostered success coping (Mehl & Pennebaker, 2003b). The EAR has also been used to obtain a measure of the frequency of different behaviors (e.g., talking on the telephone), which were then used as objective criteria for comparing the relative accuracy of self-ratings and other-ratings of behavior (Vazire & Mehl, 2008). (Close others were often as accurate as the self, although these two perspectives were often independent.)

Activity Monitoring

Accelerometers are small devices used for detecting acceleration and changes in gravity-related forces (recent wireless versions are called *wockets*). They are probably most familiar to social psychologists in iPhones and iPods, but researchers can also use them for sensing movement and activity patterns. Some researchers use accelerometers to provide objective accounts of sedentari-ness. For example, TV-watching was inversely related to general activity levels in one study (Hager, 2006), and in another, autonomous motivation for exercise predicted the frequency of moderate-intensity exercise (Standage, Sebire, & Loney, 2008). Accelerometers are also popular in sleep research. For example, the Actigraph is a relatively inexpensive wristwatch-like sensor that identifies and stores objective information about physical motion, yielding data that is highly correlated with more expensive and intrusive sleep lab polysomnography (deSouza et al., 2003). Accelerometer readings can also be used with activity recognition algorithms to identify unique motion-activity patterns, such as walking, eating, working on a computer, gesturing, and talking on the telephone, which, once identified, might generate a signal to participants to record event-contingent ratings about their thoughts and feelings (Choudhury et al., 2008).

Although accelerometers are rare in social-personality psychology, they seem useful for examining hypotheses about the relationship between activation level and mood, or about movement and activity patterns associated with individual differences, for describing patterned responses to social stimuli (e.g., freezing or fleeing a fearful stimulus, orienting one's body toward or away from a potential interaction partner), or for determining whether social interaction partners synchronize their movement. They are also likely to be helpful in applications of social psychological theories to health, where objective accounts of activity levels are desired.

Location Mapping

Global positioning systems (GPS) have become highly precise, capable of identifying a person's location within a foot or so. Moreover, this technology is readily available. Pentland notes that "the majority of adults already carry a microphone and location sensor in the form of a mobile phone, and that these sensors are packaged with computational horsepower similar to that found in desktop computers" (2007, p. 59). Location can be informative for social-psychological research. For example, location readings might be used to identify behavior settings of research interest, such as schools, work, or nature, which,

when entered, might trigger requests for self-reports of thought or affect. Intille (2007) refers to this as *context-sensitive EMA*. Another intriguing possibility would use location sensors to keep track of social network members' proximity to one another. Proximity creates opportunities for interaction, a venerable topic in interpersonal attraction research, but as yet no studies have examined systematically how physical presence leads to interaction. For example, family members might each carry with them a small badge containing a sensor that would continuously transmit location information to a central computer. These records could be combined to describe proximity among family members, co-workers, friends, or caregivers and care recipients. Continuous real-time records of this sort are ideally suited for data-intensive analytic methods, such as dynamic modeling of social influence processes (Mason, Conrey, & Smith, 2007).

TRACE MEASURES

Some social behaviors, attitudes, cognitions, and emotions leave physical traces in their wake. Bumper stickers on cars, political buttons pinned to overcoats, and posters of icons pinned to bedroom walls are all used to convey elements of attitudes, values, and identity to others. The fact that some phenomena leave residue in the physical environment raises the possibility of assessing psychological phenomena by examining the physical traces they produce.

Perhaps the most ambitious and wide-reaching effort to understand behavior from physical traces was William Rathje's garbage project. Rathje reasoned that just as archeologists use ancient refuse to learn about the behavior of people who lived many millennia ago, he too could use modern-day garbage to get insight into contemporary behavior. Thus, in 1973 he founded the garbage project at the University of Arizona with the goal of using refuse to understand contemporary patterns of consumption. In contrast to traditional studies, which had relied on questionnaires, surveys, government documents, or industry records, the garbage project was grounded in hands-on sorting of quantifiable bits and pieces of garbage. Instead of self-reports, the "garbologists" made inferences about consumer behavior directly from the material realities people left outside their houses. The investigators often found discrepancies between the answers given in self-reports and those provided by their refuse analyses. For example, in one study, "front door" interview data suggested beer consumption occurred in only 15% of the homes and was no higher than eight cans per week, whereas "back door" garbage analyses revealed that beer was consumed in 77%

of the homes, with 54% of them exceeding the supposed maximum of eight cans (Rathje & Hughes, 1975). In addition to fresh sorts of garbage bags left outside houses, the garbage project researchers also examined other sources, such as “core samples” drilled out from deep inside landfills.

At a very broad level, all trace measures rely on the processes of either erosion or accretion (Webb et al., 1981). A classic and widely cited example of erosion came from staff at the Chicago Museum of Science and Industry, who noticed that the floor tiles in front of the hatching chicks exhibit had to be replaced more frequently than those in front of other exhibits, providing an index of the relative popularity of different exhibits. Staying in the museum context, Webb et al. suggest that accretion measures too could be used to track the popularity of exhibits with glass fronts by counting the numbers of nose-prints on the glass, even making estimates of the ages of the viewers from the heights of the prints.

Building on this tradition, the personal environments that individuals craft around themselves, such as offices and bedrooms, could be rich with information about the occupants (Gosling et al., 2002). It seems likely, for example, that the pictures a person selects to hang on her walls, the books she chooses to read, and the way she arranges the items that fill the space around her all reflect aspects of her attitudes, behaviors, values, and self-views. Three different mechanisms can be delineated by which people can have an impact on the environments around them and, in turn, how physical environments can serve as repositories of individual expression (Gosling et al., 2002; Gosling, Gaddis, & Vazire, 2008). Broadly, people alter their spaces for three reasons: They want to affect how they think and feel, they want to broadcast information about themselves, and they inadvertently affect their spaces in the course of their everyday behaviors.

- *Thought and Feeling Regulators.* Personal environments are the contexts for a wide range of activities, ranging from relaxing and reminiscing to working and playing. The effectiveness with which these activities can be accomplished may be affected by the physical and ambient qualities of the space. It can be hard to relax with a lot of noise around and it is difficult to concentrate when surrounded by distractions. Specific memories, thoughts, and emotions can be evoked by mementos and photos of people, pets, and places. As a result, many items within an environment owe their presence to their ability to affect the feelings and thoughts of the occupant. Elements used to regulate emotions and thoughts could include the music on an iPod (e.g., upbeat music to get a person pumped up for a night on the town), keepsakes

on the windowsill (e.g., a twig from a tree once planted with an uncle who has since passed away), and photos of family on the refrigerator (e.g., images of an absent grandparent to evoke feelings of belonging and security).

- *Identity Claims.* One of the ways in which people make spaces their own is by adorning them with “identity claims”—deliberate symbolic statements about how they would like to be regarded (Baumeister, 1982; Swann, 1987; Swann, Rentfrow, & Guinn, 2003). Posters, awards, photos, trinkets, and other mementos are often displayed in the service of making such statements. Such signals can be split into two broad categories: Self-directed identity claims are symbolic statements made by occupants for their own benefit, intended to reinforce their self-views (e.g., displaying a fountain pen awarded in a high-school science fair). Other-directed identity claims are symbolic statements (e.g., displaying a poster of Malcolm X) about attitudes and values made to others about how one would like to be regarded.

Identity claims consist of things individuals do deliberately to their spaces, even if the occupants do not direct conscious attention to the psychological goals underlying their actions; thus, even if taping a humorous article from the satirical newspaper, *The Onion*, to one’s office door is driven by the goal of projecting a quirky nerdy cynical persona to others, it is likely that the occupant will experience the motive as “I just thought it was funny.” Of course, some identity claims are made deliberately, but that does not mean they are disingenuous; self-verification theory suggests that people make many of these claims not to create a false impression but to induce others to see them as they genuinely see themselves (Swann, this volume; Swann et al., 2002). Nonetheless, it is still possible that some claims are made with the explicit intention of fooling others (e.g., falsely claiming to admire a rock band with street credibility by wearing the band’s logo on a t-shirt). Of course, there are numerous obstacles to pulling off a successful ruse (Gosling, 2008).

- *Behavioral Residue.* Many behaviors leave some kind of discernible residue in their wake. Given that large quantities of behavior occur in personal environments, it is reasonable to suppose that these environments might accumulate a fair amount of residue. Interior behavioral residue refers to the physical traces in an environment of activities conducted within that space (e.g., an organized desk). Exterior behavioral residue refers to remnants of past activities and material preparations for activities that will take place elsewhere (e.g., a snowboard).

The elements in people's spaces are psychologically interesting phenomena in their own right but they can also be used to measure occupants' behaviors, attitudes, values, goals, and self-views. For example, cohabiting couples may use jointly acquired objects to signal things to others about their couple identity (e.g., prominently displayed honeymoon photos) or to remind themselves of special moments together (e.g., pebble from a beach where they had their first kiss); as a result, these objects may reflect the couples' relationship closeness, commitment, and dyadic adjustment (Arriaga, Goodfriend, & Lohmann, 2004; Lohmann, Arriaga, & Goodfriend, 2003). To date, only a few measures of physical spaces have been developed (Gosling, Craik, Martin, & Pryor, 2005a, 2005b). As a result, environmental evidence of social psychological behaviors has remained largely untapped despite interest in the topic in the 1960s and 1970s.

Nonetheless, the potential value of trace measures to social psychologists is great, especially given that the environmental manifestations of attitudes, values, and self-views extend well beyond physical environments. Many kinds of environments other than physical spaces (and the possessions that fill them) could furnish information about people. Just as people craft their physical spaces, they also select and mould their auditory and social environments (Mehl, Gosling, & Pennebaker, 2006; Rentfrow & Gosling, 2003, 2006). Just as people physically dwell in houses and offices, they dwell virtually in online environments like virtual worlds, personal websites, and social-networking portals (e.g., Facebook.com; Back, Schmukle, & Egloff, 2008; Vazire & Gosling, 2004). Just as people leave traces of their actions, intentions, and values in their permanent spaces, they also leave traces in other immediate surroundings such as their cars (e.g., dings in the door, unpaid scrunched-up parking tickets in the foot well, bumper stickers) or clothing (e.g., muddy running shoes, mismatched socks, a t-shirt or button with a rock band or political icon on it; Alpers & Gerdes, 2006; Vazire, Naumann, Rentfrow, & Gosling, 2008).

Thus, many environments may be used to obtain information about people. Gosling et al.'s (2002) model was developed in the context of two studies of physical environments but it can easily be applied more widely. For example, the mechanisms linking individuals to their environments can be applied to physical appearance—hairstyle and clothing can reflect identity claims, clothing and accessories can provide evidence of past or anticipated behaviors, or even levels of sexual motivation (Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007). In the domain of personality, narcissism can be expressed in terms of the kinds of clothes that people wear (e.g., expensive, stylish), their condition (e.g., organized

and neat appearance), and other malleable elements of appearance (e.g., in female targets, plucked eyebrows and cleavage showing; Vazire et al., 2008). In other words, physical appearance often holds clues to an individual's attitudes, values, intentions, behaviors, and self-views.

One study of the links between human territoriality and aggression relied on trace measures of territoriality found on cars (Szlemko, Benfield, Bell, Deffenbacher, & Troup, 2008). Starting with a definition of territoriality as a set of behaviors and cognitions that a person exhibits based on perceived ownership of space, bumper stickers, window decals, and other forms personalization served as an index of territoriality. As predicted, drivers of cars with territoriality markers scored higher on tests of driver aggression and lower on the use of constructive expressions of anger behind the wheel.

As with all methods, trace measures have their own advantages and disadvantages. One drawback is that it may be difficult to know who was responsible for a particular trace or whether the action presumed to be responsible for the trace actually caused it. For example, it may be difficult to tell whether the current or previous owner placed a bumper sticker on a car. As with any measure, the onus is on the researcher to establish its construct validity. Thus, the study of territoriality markers in cars validated the car owners' reports of bumper stickers, window decals, and so on with codings by judges made from photographs of participants' cars; the investigators also demonstrated that the presence of markers showed expected patterns of correlations with other variables such as the condition of the vehicle and the owner's attachment to it (Szlemko et al., 2008).

Past research can be used to validate trace measures. For example, research on "social snacks" supports the idea that pictures of loved ones kept in one's wallet or sitting on one's desk are used as emotion regulation strategies buffering the pain of social isolation. In one study, participants were assigned to bring to the lab either a photo of a friend or a photo of a favorite celebrity (Gardner, Pickett, & Knowles, 2005). Participants put the photos on the desks in front of them and were then asked to recall in vivid detail an experience of being rejected by other people. Unlike the people who had a picture of a celebrity in front of them, the people who were looking at a friend's image did not experience a drop in mood.

The validity of behavioral residue may also vary within a category. For example, some music genres are more tightly associated than other genres with particular values, self-views, preferences, and activities. For example, the stereotype that contemporary religious music fans place high importance on values like forgiveness, inner harmony, love, and salvation shows some accuracy, but the stereotype

that rap fans place low importance on values like a world of beauty, inner harmony, intellect, and wisdom has little accuracy (Rentfrow & Gosling, 2007). Such findings would inform researchers who use music-preference information (e.g., from iPods, CD collections) as indicators of values held by participants.

Findings that converge across methods are particularly valuable because they both underscore the robustness of the findings and cross-validate the methods. One study found converging evidence for the psychological underpinnings of political orientation by gathering data based on self-views, behavioral codings of social interactions, and records of behavioral residue (Carney, Jost, Gosling, & Potter, 2008). In particular, liberals' tendencies to be open-minded, creative, and interested in novelty seeking was reflected in high self-ratings on openness, in their tendency to smile and to be expressive and engaged in social interactions, and for their bedrooms to contain a wide variety of books (including books on travel and feminism), music (including world and classical genres), art supplies, and cultural memorabilia. Conservatives' need for order and conventionality was reflected in high self-ratings on conscientiousness and low ratings on openness, in their tendency to be detached and disengaged in social interactions, and for their bedrooms to contain organizational items (e.g., event calendars), conventional décor (e.g., sports paraphernalia, American flags), and be generally neat, clean, and organized.

One major advantage of studying behavioral residue rather than behavior itself is that it overcomes some of the significant practical challenges associated with observing behavior in natural settings (Barker, 1968; Barker & Wright, 1951; Craik, 2000; Gosling, John, Craik, & Robins, 1998; Hektner et al., 2007; Mehl et al., 2006). Moreover, whereas self-reports of behavior may underestimate actual behavioral occurrences, the existence of behavioral residue (e.g., a beer can in the trash) is usually a good sign that the behavior actually occurred. A final major benefit of residue is the advantage of aggregation. A single behavior is less reliable than a behavioral trend and physical spaces reflect behavioral trends (Epstein, 1983). For example, whereas even a generally organized person may occasionally fail to return a CD to its case and file it in the right slot, it is unlikely that such a person would have a chaotic CD collection, because a disorganized collection of CDs is the result of repeatedly engaging in similar actions.

CONCLUSION

In this chapter we have tried to describe the rationale for conducting social-psychological research outside of the

laboratory, emphasizing what it offers for the field while at the same time acknowledging its limitations. Whatever one's preferences for working inside or outside of the laboratory, we hope it is apparent that we see non-laboratory studies as neither more nor less desirable than laboratory work. Just as an artist or a craftsman uses different tools to carry out different parts of a creative work, laboratory and non-laboratory settings can provide social psychologists with different, and if used appropriately, complementary tools for our creative work. Both are intended to give researchers useful instruments for testing important theories and hypotheses about social behavior. And more important than the particular methods outlined here, we hope this chapter will serve to stimulate researchers to remain vigilant for new opportunities to examine social psychological phenomena in their natural habitats.

Most commentators agree in principle that the most valid theories and findings are those that have been tested with multiple methods in diverse settings, as noted in the introduction to this chapter. Nevertheless, current social psychological practice (as, we hasten to note, in many other disciplines) often falls short of this lofty standard. Instead, researchers tend to stick with an established paradigm, more often conducted in the laboratory with college students than anywhere else. Extending a laboratory paradigm to non-laboratory settings may sometimes require greater effort and time than conducting an additional laboratory replication but researchers who step outside the laboratory are often rewarded with increased validity and generalizability of their findings.

Kurt Lewin's call for action-oriented, real-world-relevant research, now well-aged more than a half-century, still inspires many social psychologists. Were Lewin still alive, we think he would be even more enthusiastic today about the prospects for conducting rigorous, theoretically informative and practically useful research outside of the laboratory. As we have tried to illustrate, the tools available for such research are far more sophisticated today than they were in Lewin's era. The Internet affords unparalleled access to large and diverse samples and databases. Advances in computerization, miniaturization, and cellular technology have spawned devices capable of providing extensively detailed accounts of behavior, from internal biological events to subjective states and affects to descriptions of the person's environment. Statistical methods to take advantage of these data and yield finer insights are becoming ever more sophisticated. In other words, advancing technology has made the non-laboratory environment an increasingly viable and fertile site for the generation of social-psychological knowledge. There is little doubt that these technological advances will continue and likely accelerate. As they do they will enhance our

prospects for asking and answering interesting and important questions in social psychology and beyond.

REFERENCES

- Ajilore, O., Stickgold, R., Rittenhouse, C. D., & Hobson, J. A. (1995). Nightcap: Laboratory and home-based evaluation of a portable sleep monitor. *Psychophysiology*, 32, 92–98.
- Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*, 14, 64–68.
- Alpers, G. W., & Gerdes, A. B. M. (2006). Another look at “look-alikes”: Can judges match belongings with their owners? *Journal of Individual Differences*, 27, 38–41.
- Altman, I., & Haythorn, W. W. (1965). Interpersonal exchange in isolation. *Sociometry*, 28, 411–426.
- Amir, Y. (1969). Contact hypothesis in ethnic relations. *Psychological Bulletin*, 71, 319–342.
- Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112.
- Aronson, E. (2004). Reducing hostility and building compassion: Lessons from the jigsaw classroom. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 469–488). New York: Guilford Press.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., vol. 1, pp. 1–79). Reading, MA: Addison-Wesley.
- Aronson, E., Wilson, T. D., & Brewer, M. B. (1998). Experimentation in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., vol. 1, pp. 99–142). New York: McGraw-Hill.
- Arriaga, X. B., Goodfriend, W., & Lohmann, A. (2004). Beyond the individual: Concomitants of closeness in the social and physical environment. In D. J. Mashek & A. Aron (Eds.), *Handbook of closeness and intimacy* (pp. 287–303). Mahwah, NJ: Erlbaum.
- Asch, S. E. (1952). *Social psychology*. New York: Prentice-Hall.
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality traits from email addresses. *Journal of Research in Personality*, 42, 1116–1122.
- Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138–159). New York: Cambridge University Press.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.
- Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.
- Barker, R. G., & Wright, H. F. (1951). *One boy's day. A specimen record of behavior*. New York: Harper & Row.
- Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19, 175–185.
- Bartholomew, K., Henderson, A. J. Z., & Marcia, J. E. (2000). Coded semi-structured interviews in social psychological research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 286–312). New York: Cambridge University Press.
- Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, 91, 3–26.
- 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.
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37, 245–257.
- Bevens, G. (1913). *How working men spend their time*. New York: Columbia University Press.
- Birnbaum, G. E., Reis, H. T., & Mikulincer, M. (2006). When sex is more than just sex: Attachment orientations, sexual experience, and relationship quality. *Journal of Personality and Social Psychology*, 91, 929–943.
- Birnbaum, M. H. (2001). *Introduction to behavioral research on the Internet*. Upper Saddle River, NJ: Prentice Hall.
- Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, 55, 803–832.
- Birnbaum, M. H. (in press). An overview of major techniques of web-based research. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for behavioral research on the Internet*. Washington, DC: American Psychological Association.
- Blascovich, J. (2000). Psychophysiological methods. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 117–137). New York: Cambridge University Press.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.
- Bolger, N., & Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality*, 59, 355–386.
- Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, 69, 890–902.
- Bolger, N., Zuckerman, A., & Kessler, R. C. (2000). Invisible support and adjustment to stress. *Journal of Personality and Social Psychology*, 79, 953–961.
- Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*, (pp. 3–16). New York: Cambridge University Press.
- Brondolo, E., Karlin, W., Alexander, K., Bobrow, A., & Schwartz, J. (1999). Workday communication and ambulatory blood pressure: Implications for the reactivity hypothesis. *Psychophysiology*, 36, 86–94.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. (2nd ed.). Berkeley: University of California Press.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Burleson, M., Trevathan, W. R., & Todd, M. (2007). In the mood for love, or vice versa? Understanding the relations among physical affection, sexual activity, mood, and stress in the daily lives of mid-aged women. *Archives of Sexual Behavior*, 36, 357–368.
- Burns, T. (1954). The directions of activity and communication in a departmental executive group: A quantitative study in a British engineering factory with a self-recording technique. *Human Relations*, 7, 73–97.
- Bushman, B. J. (1988). The effects of apparel on compliance: A field experiment with a female authority figure. *Personality and Social Psychology Bulletin*, 14, 459–467.
- Buunk, B. P., & Peeters, M. C. W. (1994). Stress at work, social support and companionship: Towards an event-contingent recording approach. *Work & Stress*, 8, 177–190.
- Campbell, C. G., & Molenaar, P. C. M. (in press). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*.

- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297–312.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- 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.
- Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., et al. (2008). The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7, 32–41.
- Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, 4, 53–78.
- Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, 4, 5–6.
- Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, 34, 366–375.
- Clark, L. A., Watson, D., & Leeka, J. (1989). Diurnal variation in the positive affects. *Motivation and Emotion*, 13, 205–234.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic indicators of psychological change after September 11, 2001. *Psychological Science*, 15, 687–693.
- Collins, R. L., Kashdan, T. B., & Gollnisch, G. (2003). The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. *Experimental and Clinical Psychopharmacology*, 11, 73–78.
- Conner, M. T., Perugini, M., O'Gorman, R., Ayres, K., & Prestwich, A. (2007). Relations between implicit and explicit measures of attitudes and measures of behavior: Evidence of moderation by individual difference variables. *Personality and Social Psychology Bulletin*, 33, 1727–1740.
- Conner, T., & Barrett, L. F. (2005). Implicit self-attitudes predict spontaneous affect in daily life. *Emotion*, 5, 476–488.
- Conner, T., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robins, R. C. Fraley, & R. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pp. 79–98). New York: Guilford.
- Conti, R. (2000). *Self-recording of healthy behavior*. Unpublished manuscript, Colgate University.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand-McNally.
- Cook, T. D., & Groom, C. (2004). The methodological assumptions of social psychology: The mutual dependence of substantive theory and method choice. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 19–44). Thousand Oaks, CA: Sage Publications.
- Craik, K. H. (2000). The lived day of an individual: A person-environment perspective. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-environment psychology: New directions and perspectives* (pp. 233–266). Mahwah, NJ: Erlbaum.
- Csikszentmihalyi, M., Larson, R. W., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, 6, 281–294.
- Davidovitz, R., Mikulincer, M., Shaver, P. R., Izsak, R., & Popper, M. (2007). Leaders as attachment figures: Leaders' attachment orientations predict leadership-related mental representations and followers' performance and mental health. *Journal of Personality and Social Psychology*, 93, 632–650.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (under review). *Mix me a list: Repeated presentation may not be sufficient for the Truth Effect and the Mere Exposure Effect*. Manuscript under review.
- Delespaul, P. A. E. G. (1992). Technical note: Devices and time-sampling procedures. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 363–373). Cambridge, UK: Cambridge University Press.
- deSouza, L., Benedito-Silva, A. A., Nogueira Pires, M. L., Poyares, D., Tufik, S., & Calil, M. H. (2003). Further validation of actigraphy for sleep studies. *Sleep*, 26, 81–85.
- deVries, M. W. (1992). The experience of psychopathology in natural settings: Introduction and illustration of variables. In M. W. deVries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 3–26). New York: Cambridge University Press.
- DeYoung, C. G., Hasher, L., Djikic, M., Criger, B., & Peterson, J. B. (2007). Morning people are stable people: Circadian rhythm and the higher-order factors of the Big Five. *Personality and Individual Differences*, 43, 267–276.
- Downey, G., Freitas, A. L., Michaelis, B., & Khouri, H. (1998). The self-fulfilling prophecy in close relationships: Rejection sensitivity and rejection by romantic partners. *Journal of Personality and Social Psychology*, 75, 545–560.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51, 360–392.
- Ewart, C., & Kolodner, K. B. (1994). Negative affect, gender, and expressive style predict elevated ambulatory blood pressure in adolescents. *Journal of Personality and Social Psychology*, 66, 596–605.
- Fahrenberg, J., & Myrtek, M. (Eds.). (2001). *Progress in ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Kirkland, WA: Hogrefe and Huber.
- Ferguson, E., & Chandler, S. (2005). A stage model of blood donor behaviour: Assessing volunteer behaviour. *Journal of Health Psychology*, 10, 359–372.
- Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When prophecy fails*. Minneapolis, MN: University of Minnesota Press.
- Finkel, E. J., & Eastwick, P. W. (2008). Speed-dating. *Current Directions in Psychological Science*, 17, 193–197.
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13, 83–87.
- Fraley, R. C. (2004). *How to conduct behavioral research over the Internet: A beginner's guide to HTML and CGI/Perl*. New York: Guilford.
- Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review*, 112, 60–74.
- Fraley, R. C., & Shaver, P. R. (1998). Airport separations: A naturalistic study of adult attachment dynamics in separating couples. *Journal of Personality and Social Psychology*, 75, 1198–1212.
- Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, 40, 21–34.
- Gable, S. L., & Reis, H. T. (1999). Now and then, them and us, this and that: Studying relationships across time, partner, context, and person. *Personal Relationships*, 6, 415–432.
- Gable, S. L., & Reis, H. T. (2001). Appetitive and aversive social interaction. In J. H. Harvey & A. E. Wenzel (Eds.), *Close romantic relationship maintenance and enhancement* (pp. 169–194). Mahwah, NJ: Erlbaum.
- Gable, S. L., Reis, H. T., & Downey, G. (2003). He said, she said: A quasi-signal detection analysis of daily interactions between close relationship partners. *Psychological Science*, 14, 100–105.

110 Social Psychological Methods Outside the Laboratory

- Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, 78, 1135–1149.
- Gaertner, J., Elsnor, F., Pollmann-Dahmen, K., Radbruch, L., & Sabatowski, R. (2004). *Journal of Pain and Symptom Management*, 28, 259–267.
- Gardner, W. L., Pickett, C. L., & Knowles, M. L. (2005). Social “snacking” and social “shielding”: The satisfaction of belonging needs through the use of social symbols and the social self. In K. Williams, J. Forgas, & W. von Hippel (Eds.), *The social outcast: Ostracism, social exclusion, rejection, and bullying*. New York: Psychology Press.
- Glanz, K., Murphy, S., Moylan, J., Evensen, D., & Curb, J. D. (2006). Improving dietary self-monitoring and adherence with hand-held computers: A pilot study. *American Journal of Health Promotion*, 20, 165–170.
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the Internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58, 177–193.
- Goodwin, M. S., Velicer, W. F., & Intille, S. S. (2008). Telemetric monitoring in the behavior sciences. *Behavior Research Methods*, 40, 328–341.
- Gosling, S. D. (2008). *Snoop: What your stuff says about you*. New York: Basic books.
- Gosling, S. D., & Bonnenburg, A. V. (1998). An integrative approach to personality research in anthrozoology: Ratings of six species of pets and their owners. *Anthrozoös*, 11, 184–156.
- Gosling, S. D., Craik, K. H., Martin, N. R., & Pryor, M. R. (2005a). Material attributes of Personal Living Spaces. *Home Cultures*, 2, 51–88.
- Gosling, S. D., Craik, K. H., Martin, N. R., & Pryor, M. R. (2005b). The Personal Living Space Cue Inventory: An analysis and evaluation. *Environment and Behavior*, 37, 683–705.
- Gosling, S. D., Gaddis, S., & Vazire, S. (2008). First impressions based on the environments we create and inhabit. In N. Ambady, & J. J. Skowronski (Eds.), *First Impressions* (pp. 334–356). New York: Guilford.
- Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337–1349.
- Gosling, S. D., & Johnson, J. A. (in press) *Advanced methods for behavioral research on the Internet*. Washington, DC: American Psychological Association.
- 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.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59, 93–104.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87–105.
- Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New York: Wiley-Interscience.
- Hager, R. L. (2006). Television viewing and physical activity in children. *Journal of Adolescent Health*, 39, 656–661.
- Hammond, K. R. (1998). *Ecological validity: Then and now*. Retrieved February 11, 2009, from <http://www.albany.edu/cpr/brunswick/notes/essay2.html>
- Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A., & Frederick, D. A. (2007). Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior*, 51, 40–45.
- Haslam, S. A. & McGarty, C. (2004). In C. Sansone, C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 237–264). Thousand Oaks, CA: Sage.
- Hasler, B., Mehl, M. R., Bootzin, R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, 42, 1537–1546.
- Hastie, R., & Stasser, G. (2000). Computer simulation methods for social psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 85–114). New York: Cambridge University Press.
- Hawkey, L. C., Burleson, M. H., Berntson, G. G., & Cacioppo, J. T. (2003). Loneliness in everyday life: Cardiovascular activity, psychosocial context, and health behaviors. *Journal of Personality and Social Psychology*, 85, 105–120.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage Publications.
- Helmreich, R. (1975). Applied social psychology: The unfulfilled promise. *Personality and Social Psychology Bulletin*, 1, 548–560.
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology’s view of the nature of prejudice. *Psychological Inquiry*, 19, 49–71.
- Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 11, 213–218.
- House, J. S. (1977). The three faces of social psychology. *Sociometry*, 40, 161–177.
- Hufford, M. R. (2007). Special methodological challenges and opportunities in Ecological Momentary Assessment. In A. S. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 54–75). New York: Oxford University Press.
- Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors*, 16, 205–211.
- Ickes, W. (1983). A basic paradigm for the study of unstructured dyadic interaction. *New Directions for Methodology of Social & Behavioral Science*, 15, 5–21.
- Intille, S. S. (2007). Technological innovations enabling automatic, context-sensitive ecological momentary assessment. In A. S. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 308–337). New York: Oxford University Press.
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995–1006.
- Jensen-Campbell, L. A., & Graziano, W. G. (2000). Beyond the school yard: Relationships as moderators of daily interpersonal conflict. *Personality and Social Psychology Bulletin*, 26, 923–935.
- Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, 19, 15–29.
- Johnson, J. A. (2005). Ascertaining the validity of web-based personality inventories. *Journal of Research in Personality*, 39, 103–129.
- Joinson, A. N., McKenna, K., Postmes, T., & Reips, U.-D. (Eds.) (2007). *The Oxford handbook of Internet psychology*. Oxford, UK: Oxford University Press.
- Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., vol. 1, pp. 47–107). New York: Random House.
- Josephson, W. L. (1987). Television violence and children’s aggression: Testing the priming, social script, and disinhibition predictions. *Journal of Personality and Social Psychology*, 53, 882–890.

- Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology*, 43, 1–23.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (vol. 1, 4th ed., pp. 233–265). New York: McGraw-Hill.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.
- Kerig, P. K., & Lindahl, K. M. (Eds.). (2001). *Family observational coding systems: Resources for systemic research*. Mahwah, NJ: Erlbaum.
- Kirby, K. C., Fowler, S. A., & Baer, D. M. (1991). Reactivity in self-recording: Otrusiveness of recording procedure and peer comments. *Journal of Applied Behavior Analysis*, 24, 487–498.
- Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, 11, 415–425.
- Krantz, J. H. (2001). Stimulus delivery on the Web: What can be presented when calibration isn't possible? In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 113–130). Lengerich, Germany: Pabst Science Publishers.
- Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35–60). San Diego, CA: Academic Press.
- Krantz, J. H., & Williams, J. E. (in press). Using media for Web-based research. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for behavioral research on the Internet*. Washington, DC: American Psychological Association.
- Krosnick, J. A., & Fabrigar, L. R. (in press). *The handbook of questionnaire design*. New York: Oxford University Press.
- Lane, R., Reis, H. T., Peterson, D., Zareba, W., & Moss, A. (2008). *Emotion in Long QT Syndrome*. Manuscript in preparation.
- Langer, E. J., & Rodin, J. (1976). The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology*, 34, 191–198.
- Larsen, R. J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology*, 52, 1195–1204.
- Larsen, R. J., & Kasimatis, M. (1991). Day-to-day physical symptoms: Individual differences in the occurrence, duration, and emotional concomitants of minor daily illnesses. *Journal of Personality*, 59, 387–423.
- Larson, R. W., Richards, M. H., & Perry-Jenkins, M. (1994). Divergent worlds: The daily emotional experience of mothers and fathers in the domestic and public spheres. *Journal of Personality and Social Psychology*, 67, 1034–1046.
- Laurenceau, J.-P., & Bolger, N. (2005). Using diary methods to study marital and family processes. *Journal of Family Psychology*, 19, 86–97.
- Lebo, H. (2000). *The UCLA Internet report: Surveying the digital future*. Retrieved July 1, 2004, from <http://www.ccp.ucla.edu/UCLAInternet-Report-2000.pdf>.
- Lenhart, A. (2000). *Who's not online: 57% of those without Internet access say they do not plan to log on*. Retrieved July 1, 2004, from <http://www.pewinternet.org/reports/toc.asp?Report=21>
- Levav, J., & Fitzsimons, G. J. (2006). When questions change behavior: The role of ease of representation. *Psychological Science*, 17, 207–213.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper.
- Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology*, 17, 48–52.
- Lohmann, A., Arriaga, X. B., & Goodfriend, W. (2003). Close relationships and placemaking: Do objects in a couple's home reflect couplehood? *Personal Relationships*, 10, 439–451.
- Luce, K. H., Winzelberg, A. J., Das, S., Osborne, M. I., Bryson, S. W., & Taylor, C. B. (2007). Reliability of self-report: Paper versus online administration. *Computers in Human Behavior*, 23, 1384–1389.
- Mahoney, M. J., Moura, N. G., & Wade, T. C. (1973). Relative efficacy of self-reward, self-punishment, and self-monitoring techniques for weight loss. *Journal of Consulting and Clinical Psychology*, 40, 404–407.
- Mangan, M., & Reips, U.-D. (2007). Sleep, sex, and the Web: Surveying the difficult-to-reach clinical population suffering from sexomnia. *Behavior Research Methods*, 39, 233–236.
- Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal websites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, 90, 1014–1031.
- Mason, W. A., Conrey, F. R., & Smith, E. R. (2007). Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. *Personality and Social Psychology Review*, 11, 279–300.
- Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat: Spontaneous expressions of medal winners of the 2004 Athens Olympic games. *Journal of Personality and Social Psychology*, 91, 568–581.
- McClelland, D. C. (1957). Toward a science of personality psychology. In H. P. David & H. von Bracken (Eds.), *Perspective in personality theory* (pp. 355–382). New York: Basic Books.
- McGrath, J. E., & Altermatt, T. W. (2000). Observation and analysis of group interaction over time: Some methodological and strategic choices. In M. Hogg & R. S. Tindale (Eds.), *Blackwell's handbook of social psychology: Vol. 3. Group processes* (pp. 525–556). London: Blackwell Publishers.
- McGuire, W. J. (1967). Some impending reorientations in social psychology: Some thoughts provoked by Kenneth Ring. *Journal of Experimental Social Psychology*, 3, 124–139.
- 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., & Pennebaker, J. W. (2003a). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84, 857–870.
- Mehl, M. R., & Pennebaker, J. W. (2003b). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science*, 14, 579–585.
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorded (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments & Computers*, 33, 517–523.
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slater, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82.
- Merrilees, C. E., Goeke-Morey, M., & Cummings, E. M. (2008). Do event-contingent diaries about marital conflict change marital interactions? *Behaviour Research and Therapy*, 46, 253–262.
- Milgram, S., Mann, L., & Harter, S. (1965). The lost-letter technique: A tool of social research. *Public Opinion Quarterly*, 29, 437.
- Miller, G., Tybur, J. M., & Jordan, B. D. (2007). Ovulatory cycle effects on tip earnings by lap dancers: Economic evidence for human estrus? *Evolution and Human Behavior*, 28, 375–381.
- Moghaddam, N. G., & Ferguson, E. (2007). Smoking, mood regulation, and personality: An event-sampling exploration of potential models and moderation. *Journal of Personality*, 75, 451–478.
- Mohr, C. D., Armeli, S., Tennen, H., Carney, M. A., Affleck, G., & Hromi, A. (2001). Daily interpersonal experiences, context, and alcohol consumption: Crying in your beer and toasting good times. *Journal of Personality and Social Psychology*, 80, 489–500.

112 Social Psychological Methods Outside the Laboratory

- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387.
- Moriarty, T. (1975). Crime, commitment, and the responsive bystander: Two field experiments. *Journal of Personality and Social Psychology*, 31, 370–376.
- Murray, S. L., Griffin, D. W., Rose, P., & Bellavia, G. M. (2003). Calibrating the sociometer: The relational contingencies of self-esteem. *Journal of Personality and Social Psychology*, 85, 63–84.
- Musch, J., & Reips, U.-D. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61–88). San Diego, CA: Academic Press.
- Nelson, R. O. (1977). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler & P. M. Miller (Eds.), *Progress in behavior modification* (vol. 5, pp. 273–297). New York: Academic Press.
- Neubarth, W. (in press). Drag and drop: Using moveable objects to let participants measure and rank items. In S. D. Gosling and J. A. Johnson (Eds.), *Advanced methods for behavioral research on the Internet*. Washington, DC: American Psychological Association.
- Newcomb, T. M. (1961). *The acquaintance process*. New York: Holt, Rinehart & Winston.
- Nezlek, J. B. (2003). Using multilevel random coefficient modeling to analyze social interaction diary data. *Journal of Social and Personal Relationships*, 20, 437–469.
- Nezlek, J. B., Kafetsios, K., & Smith, C. V. (2008). Emotions in everyday social encounters: Correspondence between culture and self-construal. *Journal of Cross-Cultural Psychology*, 39, 366–372.
- North, A. C. & Hargreaves, D. J. (1997). In-store music affects product choice. *Nature*, 390, 132.
- Oishi, S., Schimmack, U., & Diener, E. (2001). Pleasures and subjective well-being. *European Journal of Personality*, 15, 153–167.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2008). *Instructional manipulation checks: Detecting satisficing to increase statistical power*. Manuscript under review.
- Patrick, H., Knee, C. R., Canevello, A., & Lonsbary, C. (2007). The role of need fulfillment in relationship functioning and well-being: A self-determination theory perspective. *Journal of Personality and Social Psychology*, 92, 434–457.
- Pemberton, M. B., Insko, C. A., & Schopler, J. (1996). Memory for and experience of differential competitive behavior of individuals and groups. *Journal of Personality and Social Psychology*, 71, 953–966.
- Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, 8, 162–166.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547.
- Pennebaker, J. W., & Newton, D. (1983). Observation of a unique event: The psychological impact of the Mount Saint Helens Volcano. *New Directions for Methodology of Social & Behavioral Science*, 15, 93–109.
- Pentland, A. (2007). Automatic mapping and modeling of human networks. *Physica A*, 378, 59–67.
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90, 751–783.
- Petty, R. E., & Cacioppo, J. T. (1996). Addressing disturbing and disturbed consumer behavior: Is it necessary to change the way we conduct behavioral science? *Journal of Marketing Research*, 33, 1–8.
- Piasecki, T. M., Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment*, 19, 25–43.
- Piliavin, I. M., Rodin, J., & Piliavin, J. A. (1969). Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology*, 13, 289–299.
- Rafaeli, E. (2009). Daily diary methods. In H. T. Reis & S. Sprecher (Eds.), *Encyclopedia of human relationships*. Thousand Oaks, CA: Sage.
- Raikkonen, K., Matthews, K. A., Flory, J. D., Owens, J. F., & Gump, B. B. (1999). Effects of optimism, pessimism, and trait anxiety on ambulatory blood pressure and mood during everyday life. *Journal of Personality and Social Psychology*, 76, 104–113.
- Rathje, W. L., & Hughes, W. W. (1975). The garbage project as a nonreactive approach: Garbage in...garbage out? In H. W. Sinaiko, & L. A. Broedling (Eds.) *Perspectives on attitude assessment: Surveys and their alternatives*. Washington: Smithsonian Institution.
- Raudenbush, S. W., Brennan, R. T., & Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology*, 9, 161–174.
- Reips, U.-D. (2000). The Web Experiment Method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89–118). San Diego, CA: Academic Press.
- Reips, U.-D. (2002a). Standards for Internet-based experimenting. *Experimental Psychology*, 49, 243–256.
- Reips, U.-D. (2002b). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review*, 20, 241–249.
- Reips, U.-D. (2002c). Theory and techniques of Web experimenting. In B. Batonic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 229–250). Seattle: Hogrefe & Huber.
- Reips, U.-D., & Krantz, J. (in press). Experimental data. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for behavioral research on the Internet*. Washington, DC: American Psychological Association.
- Reips, U.-D., & Neuhaus, C. (2002). WEXTOR: A Web-based tool for generating and visualizing experimental designs and procedures. *Behavior Research Methods, Instruments, & Computers*, 34, 234–240.
- Reis, H. T. (Ed.). (1983). *Naturalistic approaches to studying social interaction*. San Francisco: Jossey-Bass.
- Reis, H. T. (1994). Domains of experience: Investigating relationship processes from three perspectives. In R. Erber & R. Gilmour (Eds.), *Theoretical frameworks for personal relationships* (pp. 87–110). Hillsdale, NJ: Erlbaum.
- Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York: Cambridge University Press.
- Reis, H. T., Sheldon, K., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, 26, 419–435.
- Reis, H. T., & Wheeler, L. (1991). Studying social interaction with the Rochester Interaction Record. *Advances in experimental social psychology* (Vol. 24, pp. 269–318). San Diego, CA: Academic Press.
- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84, 1236–1256.
- Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: Personality judgments based on music preferences. *Psychological Science*, 17, 236–242.
- Rentfrow, P. J., & Gosling, S. D. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of Music*, 35, 306–326.
- Repetti, R. L. (1989). Effects of daily workload on subsequent behavior during marital interaction: The roles of social withdrawal and spouse support. *Journal of Personality and Social Psychology*, 57, 651–659.
- Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, 3, 113–123.

- Robinson, J. P., & Godbey, G. (1997). *Time for life: The surprising ways Americans use their time*. University Park: The Pennsylvania State University Press.
- Rodin, J., & Langer, E. J. (1977). Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of Personality and Social Psychology*, 35, 897–902.
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5, 2–14.
- Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward? *Perspectives on Psychological Science*, 4, 435–439.
- Salovey, P., & Williams-Piehota, P. (2004). Field experiments in social psychology: Message framing and the promotion of health protective behaviors. *American Behavioral Scientist*, 47, 488–505.
- Sbarra, D. A., & Emery, R. E. (2005). The emotional sequelae of non-marital relationship dissolution: Analysis of change and intraindividual variability over time. *Personal Relationships*, 12, 213–232.
- Schwartz, J. E., & Stone, A. A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology*, 17, 6–16.
- Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 143–179). New York: McGraw-Hill.
- Sears, D. O. (1986). College sophomores in the lab: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.
- 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 Book Exchange.
- Sherman, D. K., & Kim, H. S. (2005). Is there an "I" in "Team"? The role of the self in group-serving judgments. *Journal of Personality and Social Psychology*, 88, 108–120.
- Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M., & Finney, L. J. (1999). Twenty years of PSPB: Trends in content, design and analysis. *Personality and Social Psychology Bulletin*, 25, 177–187.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4, 1–32.
- Shohat, M., & Musch, J. (2003). Online auctions as a research tool: a field experiment on ethnic discrimination. *Swiss Journal of Social Psychology*, 62, 139–145.
- Shulman, A. D., & Berman, H. J. (1975). Role expectations about subjects and experimenters in psychological research. *Journal of Personality and Social Psychology*, 32, 368–380.
- Silver, R. C. (2004). Conducting research after the 9/11 terrorist attacks: Challenges and results. *Families, Systems & Health*, 22, 47–51.
- Silverman, I. (1971). Crisis in social psychology: The relevance of relevance. *American Psychologist*, 26, 583–584.
- Silvia, P. J. (2002). Self-awareness and the regulation of emotional intensity. *Self and Identity*, 1, 3–10.
- Simonton, D. K. (2003). Qualitative and quantitative analyses of historical data. *Annual Review of Psychology*, 54, 617–640.
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology*, 57, 529–555.
- Skowronski, J. J., Betz, A. L., Thompson, C. P., & Shannon, L. (1991). Social memory in everyday life: Recall of self-events and other-events. *Journal of Personality and Social Psychology*, 60, 831–843.
- Smith, E. R. (2000). Research design. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 17–39). New York: Cambridge University Press.
- Smith, E. R., & Mackie, D. M. (2000). *Social psychology* (2nd ed.). Philadelphia: Psychology Press.
- Sriram, N. (2002). The role of gender, ethnicity, and age in intergroup behavior in a naturalistic setting. *Applied Psychology: An International Review*, 51, 251–265.
- Standage, M., Sebire, S. J., & Loney, T. (2008). Does exercise motivation predict engagement in objectively assessed bouts of moderate-intensity exercise?: A self-determination theory perspective. *Journal of Sport & Exercise Psychology*, 30, 337–352.
- Stone, A. A., Broderick, J. E., Porter, L. S., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care & Research*, 10, 185–193.
- Stone, A. A., Hedges, S. M., Neale, J. M., & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a "blue Monday" phenomenon. *Journal of Personality and Social Psychology*, 49, 129–134.
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199–202.
- Stone, A. A., & Shiffman, S. (2007). Capturing momentary self-report data: A proposal for reporting guidelines. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 371–386). New York: Oxford University Press.
- Stone, A. A., Shiffman, S., Atienza, A. A., & Nebeling, L. (Eds.). (2007). *The science of real-time data capture*. New York: Oxford University Press.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, 324, 1193–1194.
- Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Erlbaum.
- Story, L. B., & Repetti, R. (2006). Daily occupational stressors and marital behavior. *Journal of Family Psychology*, 20, 690–700.
- Suls, J., Martin, R., & David, J. P. (1998). Person-environment fit and its limits: Agreeableness, neuroticism, and emotional reactivity to interpersonal conflict. *Personality and Social Psychology Bulletin*, 24, 88–98.
- Swann, W. B., Jr. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology*, 53, 1038–1051.
- Swann, W. B., Jr., Rentfrow, P. J., & Guinn, J. (2003). Self-verification: The search for coherence. In M. Leary & J. Tagney (Eds.), *Handbook of self and identity*. New York: Guilford Press.
- Swim, J. K., Hyers, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday sexism: Evidence for its incidence, nature, and psychological impact from three daily diary studies. *Journal of Social Issues*, 57, 31–53.
- Szlemko, W. J., Benfield, J. A., Bell, P. A., Deffenbacher, J. L., & Troup, L. (2008). Territorial markings as predictors of drive aggression and road rage. *Journal of Applied Social Psychology*, 38, 1665–1688.
- Takarangi, M. K., Garry, M., & Loftus, E. F. (2006). Dear diary, is plastic better than paper? I can't remember: Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006). *Psychological Methods*, 11, 119–122.
- Tennen, H., Affleck, G., Coyne, J. C., Larsen, R. J., & DeLongis, A. (2006). Paper and plastic in daily diary research: Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006). *Psychological Methods*, 11, 112–118.
- Thomas, D., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291–297.
- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, 9, 507–533.
- Turner, C. W., Layton, J. F., & Simons, L. S. (1975). Naturalistic studies of aggressive behavior: Aggressive stimuli, victim visibility, and horn honking. *Journal of Personality and Social Psychology*, 31, 1098–1107.

114 Social Psychological Methods Outside the Laboratory

- Vazire, S., & Gosling, S. D. (2004). e-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, 87, 123–132.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, 95, 1202–1216.
- Vazire, S., Naumann, L. P., Rentfrow, P. J., & Gosling, S. D. (2008). Portrait of a narcissist: Manifestations of narcissism in physical appearance. *Journal of Research in Personality*, 42, 1439–1447.
- Visser, P. S., Krosnick, J. A., & Lavrakas, P. J. (2000). Survey research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 223–252). New York: Cambridge University Press.
- Walls, T. A., Schafer, J. L. (2006) (Eds.). *Models for intensive longitudinal data*. Oxford, UK: Oxford University Press.
- Walster, E., Walster, G. W., & Lambert, P. (1971). Playing hard-to-get: A field study. Reported in Walster, E., Walster, G. W., Piliavin, J., & Schmidt, L. (1973). "Playing hard to get": Understanding an elusive phenomenon. *Journal of Personality and Social Psychology*, 26, 113–121.
- Walther, J. B., van der Heide, B., Kim, S.-Y., Westerman, D., & Tong, S. T. (2008). The role of friends' appearance and behavior on evaluations of individuals on Facebook: Are we known by the company we keep? *Human Communication Research*, 34, 28–49.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (2000). *Unobtrusive measures* (2nd ed.). Thousand Oaks, CA: Sage.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Boston: Houghton Mifflin.
- Weick, K. E. (1985). Systematic observational methods. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 567–634). New York: Random House.
- West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). New York: Cambridge University Press.
- West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, 59, 609–662.
- West, S. G., Newsom, J. T., & Fenaughty, A. M. (1992). Publication trends in JPSP: Stability and change in topics, methods, and theories across two decades. *Personality and Social Psychology Bulletin*, 18, 473–484.
- Wethington, E., & Almeida, D. M. (in press). Assessing daily well-being via telephone diaries. In R. Belli, F. P. Stafford, & D. F. Alwin (Eds.), *Using calendar and diary methods in life events research*. Thousand Oaks, CA: Sage.
- Wheeler, L., & Miyake, K. (1992). Social comparison in everyday life. *Journal of Personality and Social Psychology*, 62, 760–773.
- Wheeler, L., & Nezlek, J. (1977). Sex differences in social participation. *Journal of Personality and Social Psychology*, 35, 742–754.
- Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, 59, 339–354.
- White, W. B., Schulman, P., McCabe, E. J., & Dey, H. M. (1989). Average daily blood pressure, not office blood pressure, determines cardiac function in patients with hypertension. *JAMA*, 261, 873–877.
- Williams, K. D. (2001). *Ostracism: The power of silence*. New York: Guilford.
- Woike, B. A. (1995). Most-memorable experiences: Evidence for a link between implicit and explicit motives and social cognitive processes in everyday life. *Journal of Personality and Social Psychology*, 68, 1081–1091.
- Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, 83, 1281–1297.
- Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology and Behavior*, 10, 115–121.
- Zucker, R. A., Manosevitz, M., & Lanyon, R. I. (1968). Birth order, anxiety, and affiliation during a crisis. *Journal of Personality and Social Psychology*, 8, 354–359.