

Back to the Future: Methodologies That Capture Real People in the Real World

Ho Kwan Cheung¹, Mikki Hebl², Eden B. King², Hannah Markell¹, Carlos Moreno², and Christy Nittrouer²

Social Psychological and
Personality Science
1-9

© The Author(s) 2017
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1948550617709113
journals.sagepub.com/home/spp



Abstract

This article argues that social and personality psychology theory and findings have been unnecessarily hampered by our reliance on a set of methodological lenses that sacrifice external validity and generalizability. We further urge researchers to renew their use of both classic and evolving alternative methodologies. In particular, we discuss and consider pros and cons of field experiments, big data, archival research, and observations. We propose that it is up to us, as researchers, to ensure that we maximize the breadth of our theory and quality of data by expanding and triangulating the ways in which we collect data.

Keywords

methodology, field experiments, big data, observations, archival research

Participation in experiments has become a rite of passage for college students enrolled in introductory-level, college psychology classes. Psychology researchers know just how “convenient” the college student convenience sample is—it can be an extraordinarily inexpensive and effective way to recruit participants to test causal mechanisms underlying psychological phenomena (Druckman & Kam, 2009). However, researchers have also long recognized the limitations of sole reliance on laboratory experiments. Indeed, three decades ago, Sears (1986) highlighted two dangers affiliated with this “narrow database”: (1) our conclusions may not generalize beyond the average American college student and (2) convenience may come at the expense of external validity.

Questions of population sampling and external validity of laboratory experiments have persisted (see Henrich, Heine, & Norenzayan, 2010; Peterson, 2001) and have been applied to additional controversial subject pools, including mTurk, RapidWorkers, MicroWorkers, ClickWorker, and other crowdsourcing sites. The participant recruitment efficiency benefits mask the darker side of online or web-based subject pools such as the ambiguity of population representation, inappropriately motivated responding, survey respondents participating in multiple studies conducted by the same research laboratory, and the exploitation of participants (see Dholakia, 2017; Zhou & Fishbach, 2016). Moreover, participants in web-based samples may say they are from “X,” but their internet protocol address reveals that they are actually located in “Y” and may be motivated to complete surveys as quickly as possible rather than as honestly and completely (Zhou & Fishbach, 2016). Furthermore, Dholakia suggests that we

have moved to an overreliance on such data collection and concludes that “there are not that many fish and there are far too many nets cast into the MTurk pond.”

In response to these concerns, Henrich, Heine, and Norenzayan (2010) suggested that researchers “need to be less cavalier in addressing questions of human nature on the basis of data drawn from this particularly thin, and rather unusual, slice of humanity” (p. 61). Consequently, this article focuses on four alternative methodologies that enhance external validity and/or generalizability to varying degrees—field experiments, big data, archival data, and observations. These methodologies require researchers to get outside of their laboratories and stop recruiting participants electronically for research and therefore better examine *real* people and *real* data in the real world. This is not to say that these methodologies are a panacea; rather, and as demonstrated in Table 1, there are benefits and drawbacks to every methodology. Nevertheless, diversifying our methodology is an important step toward building a broader and more representative science of human behavior. At a time when the field is facing a crisis of confidence and problems with falsification and replicability (Pashler & Wagenmakers, 2012), one answer may lie in moving beyond the current zeitgeist in data collection and using different—be they classic or

¹ Department of Psychology, George Mason University, Fairfax, VA, USA

² Department of Psychology, Rice University, Houston, TX, USA

Corresponding Author:

Mikki Hebl, Department of Psychology, Rice University, 6100 Main Street, Houston, TX 77005, USA.

Email: hebl@rice.edu

Table 1. Underutilized Social Psychological Research Methods.

| Description | Field Experiments | Big Data Methods | Archival Data Methods | Observational Methods |
|-------------|---|--|--|---|
| Definition | “A methodology that uses random assignment to implement a manipulation relevant to working adult participants engaging in genuine tasks or with genuine outcomes in natural settings” (King, Hebl, Botsford Morgan, & Ahmad, 2013, p. 502) | Applying “powerful computational techniques to unveil trends and patterns within and between . . .” (George, Haas, & Pentland, 2014, p. 321) high-volume data sets with diverse and dynamic elements | Analyzing data collected prior to the formation of a particular study or research questions | The observation and description of subjects’ ongoing behavior |
| Benefits | <ul style="list-style-type: none"> • Precise measurement • Highly generalizable context • Creativity/originality • Allows for causal inferences • Reduces publication bias • Enhances average participant diversity • Highly naturalistic • Informs and develops accurate and consequential psychological theory • Because they generally require more resources, may be more replicable and provide more accurate measures of effect size • Practical implications | <ul style="list-style-type: none"> • Large quantities that can include entire populations • Continuous data collection and analysis • Largely unobtrusive • Data can be updated automatically and in real time • Amount of data being collected can be limitless | <ul style="list-style-type: none"> • Allows one to access unreachable populations • Can be more rigorous methodology with state or national-level random sampling • Can be cost-effective • Lots of power, large samples • Can provide longitudinal and multilevel answers • Can examine influences of individual and environmental forces | <ul style="list-style-type: none"> • Often easy access • Observations of direct behavior • Ecologically and externally valid • Rich look into social behavior and contextual variables • Can examine dangerous or unethical behavior not possible in the lab |
| Drawbacks | <ul style="list-style-type: none"> • Spontaneous problems may arise • Lack of control over all variables • Sometimes complex • Can be difficult to preserve methodological rigor when difficulties arise “on the fly” • Requires significant expense of time, resources, and energy | <ul style="list-style-type: none"> • Can be atheoretical • Ethical issues • Measurement weaknesses • Analysis challenges • Technology hasn’t caught up with ideas • Often lack control • May require setting up an entirely separate mechanism to handle the data collection and analysis • At times, coming up with research ideas to fit the data being collected instead of the reverse | <ul style="list-style-type: none"> • Might not know that such data exists • Might not know where to find archives • Might not know how to use such data • Need to consider large number of control variables due to larger degree of variability among the diverse sample • May not be a perfect match for one’s theory, question, or model | <ul style="list-style-type: none"> • Cannot draw causal inferences • Spontaneous issues may arise, making observations difficult • Observer bias • Problem of objectivity • Coding complexity and long duration |

novel—paradigms that show triangulation of results with diverse sets of methodologies and research participants. As such, we now describe and discuss four different methodologies, the pros and cons of which are summarized in Table 1.

Field Experiments

As Lewin said in 1939, “To explain social behavior it is necessary to represent the structure of the total situation and the distribution of the forces in it” (p. 868). Such a description captures the essence of the field experiment. Rather than just relying on observational data in a naturalistic setting, a field experiment is a fully randomized research design in which at

least one variable is manipulated or in which observations are assigned to treatment and control conditions. Field experiments maximize measurement precision by assessing effects in uncontrolled settings (“the wild”), enhance context fidelity, increase generalizability, reduce the likelihood of publication bias (“the tendency for researchers to publish positive results” [p. 104]) since the high expenditure of resources to conduct one study makes it unlikely that the study was accompanied by several “failed” attempts at detecting an intervention effect, and enhance average participant diversity in ways that laboratory experiments and survey methods typically do not (see also Maner, 2016). Further, field experiments often incorporate creative approaches to problem-solving and experimentation

that break the mold of the frequently template-like format of survey and lab experiments.

Unfortunately, field experiments are rarely used by social psychologists. There are exceptions, of course, and we point to the excellent work of Paluck as being particularly illustrative of how well such research can be done and the promise it offers. As Paluck, herself, states, “Social scientists have spent enough time in the lab learning about the mechanisms of interventions with no known real-world effects” (Paluck, 2016, p. 147). Most recently, then, Paluck, Shepherd, and Aronow (2016) designed an experimental intervention to examine whether social referents could influence conflict and bullying behaviors at schools. Examining 56 public middle schools across New Jersey, they found that spreading anticonflict norms was successful in reducing disciplinary events, particularly when those who were trained to be referents were highly connected students (see also Paluck, 2011; Paluck & Shepherd, 2012). This sort of modern-day study that would no doubt make Lewin happy.

In our own labs, we have run a number of field experiments, too, and particularly used these to show the manifestation of discrimination in subtle (as opposed to overt) forms. In particular, we have had research participants act the part of job candidates and/or customers in retail stores. Such participants have worn (a) hats that portray them as “Gay and Proud” (a stigmatized condition) or “Texan and Proud” (a control condition), (b) obesity prosthetics, (c) pregnancy prosthetics, and (d) hijabs (see Hebl, Foster, Mannix, & Dovidio, 2002; Hebl, King, Glick, Singletary, & Kazama, 2007; King & Ahmad, 2010; King, Shapiro, Hebl, Singletary, & Turner, 2006). Additionally, participants in many of these studies have enacted strategies and/or articulated statements that have allowed us to measure potential reductions in the discrimination that many stigmatized individuals receive. And in nearly all of these studies, we have found that evidence that subtle discrimination in job applications can be overcome through stereotype-inconsistent behaviors. Is it conceivable to run such studies in the lab or with mTurk respondents? We think not. The conduct of experimental studies in the field allowed us to, as Lewin stated, capture the situation and the forces in it.

The studies we described provide just a few examples of the strength of systematic, controlled, and high-fidelity field experiments (see Table 1). As Paluck and Cialdini (2014) comment, advantages to these types of experiments include the fact that an investigator does not have to work as hard to make the units, treatment, observations, or study naturalistic. This is not to imply that conducting rigorous field experiments is easy—it can oftentimes require a different set of skills than those required in laboratory experiments. For example, when (often not “if”) things go wrong in the field, a researcher is often forced to improvise spontaneously and come up with in-the-moment contingencies that will not sacrifice the methodological rigor of the study. Further, although one of the advantages of the field is a high-fidelity context, one of the drawbacks is a lack of control of the variables the researcher is testing. Thus, the onus is on the researcher to exert as much control as

possible over the experimental conditions, and do one’s best to think through all the threats to internal validity that may be unique to that particular field setting. This said, the advantages afforded by field experiments, as well as their power to generate accurate and highly consequential psychological theory, indicate that they are powerful and should be implemented more often.

When thinking about how to tackle the actual execution of a field experiment, coming up with a testable idea, feasible location, and the myriad of other necessary logistics can seem overwhelming—to say the least (see King, Hebl, Botsford Morgan, & Ahmad, 2013). The best way to get started with a good field experiment is for researchers to think about natural field settings to which they have access (e.g., either personally or by leveraging their networks). Then, researchers should think about starting with the variables critical for any given setting and what they would most like to manipulate. When choosing these variables, a helpful starting place is for researchers to think about what variable might have conditions that would result in the greatest degree of behavior change whether introduced into the setting. Complications may set in, but with practice, dealing with such issues become fairly routinized. And the excitement of data collection in conditions that are more realistic is exciting, meaningful, and well worth it.

Archival Data

Field experiments balance internal and external validity but rarely fully address concerns about sampling or generalizability. Instead, repositories of representative data—often free and publicly accessible—may be useful to assess the generalizability of some findings. For many reasons, however, social psychologists tend to stay away from this sort of data. This may be because social psychologists tend to be focused on collecting and analyzing new data, they do not know where to look for this sort of data or the topics to examine with it or they lack training in using such data (see also Shultz, Hoffman, & Reiter-Palmon, 2005). In the most basic sense, this secondary data are referred to as archival data, which can be defined as any data collected prior to a study. In many social science fields, such as sociology, economics, and anthropology, archival data are often used to answer complex empirical questions of human behavior and social phenomena. Even though archival data sets are not specifically designed to address most researchers’ particular theories or models, they offer many unique advantages and contain extensive information that can help social psychologists answer questions that are otherwise difficult to answer with more traditional methods.

First, archival data sets allow us to tap into samples that are normally unreachable to researchers due to limited financial and social resources. It has long been argued that our psychological science suffers from the prevalence of underpowered studies and convenience sampling (Arnett, 2008; Maxwell, 2004). Archival data, especially in the form of state or national-level data sets generated by rigorous random sampling methods, can be a solution to this problem by providing

information and insight into a large, diversity of samples that are more representative of the general population about which researchers wish to generalize their findings. This increases the statistical power and external validity of the research findings. Second, archival data can help answer questions that are either longitudinal or multilevel in nature, which can be difficult or even impossible to capture with any single research team's data collection effort. With the heavy emphasis of our field on the interaction between the environment and individual, archival data sets can be especially useful to answer questions regarding the effects of different contextual factors on individuals' behaviors. For instance, King et al. (2011) utilized four separate publicly available data sets, including the census, from the United Kingdom to examine the effects of community and organizational demography on the individual level of workplace incivility. Furthermore, archival data can also be useful in understanding changes over time. One such example is the Terman life cycle study, which was originally set up to better understand the lives of children with high intelligence by interviewing 1,528 children from schools for gifted children in California. Archival data have also been used to study various phenomena from the long-term effects of parental divorce to different methods of adaptation utilized by the elderly (Crosnoe & Elder, 2002; Martin, Friedman, & Schwartz, 2007; Tucker et al., 1997).

As people spend increasing amounts of time using social media and surfing the web, the Internet also serves as a newer source of archival data that can lend unique insights into individuals' thoughts, attitudes, and behaviors by avoiding the effect of experimenters and common biases associated with self-reported measures. With the increasing capacity of technology to collect massive amount of information, some of these archival data can be classified as big data, which we will describe more in detail in a later section. One valuable type of Internet-based archival data comes from social media (e.g., Facebook and Twitter), in which people express their opinions in a myriad of direct (e.g., commenting on someone's status, posting a photo) and indirect (e.g., liking a video, clicking to view an article) forms. Garcia and Sikström (2014) used latent semantics analysis to analyze Facebook status updates to predict personality traits, while Hernandez, Newman, and Jeon (2015) used Twitter feeds to examine variability in job satisfaction. In addition to social media, record numbers of Internet searches, such as Google search trends, can be a powerful indicator of social attitudes. As exemplified by the recent Brexit referendum in the United Kingdom, Google (2016) demonstrated a surge in searches for "European Union" after the polls closed, suggesting the public did not fully consider the implications of the referendum until after it had passed. Because the use of search engine data is still its infancy, there have been few published academic studies utilizing this method. Nevertheless, it has the potential to serve as a valuable tool for social and personality psychologists to better understand implicit attitudes and opinions.

Because archival data already are widely used by social scientists in other disciplines, there are existing platforms that already make them accessible to researchers. Specifically, in

the United States, some available social science database archives include the Interuniversity Consortium for Political and Social Research (<http://www.icpsr.umich.edu>), Murray Research Archive (<http://www.radcliffe.edu/murray>), and Families and Work Institute (<http://www.murray.harvard.edu>). Additionally, researchers can gain access to the Census Bureau and Department of Labor (see also cites listed as Shultz et al., 2005). Researchers can easily search for data sets that contain variables of interests across different subdisciplines. Many of these data sets also allow researchers to download codebooks and even perform basic descriptive statistical analyses before downloading data, so they can decide whether a given data set is the correct fit for their research question. As for more modern forms of archival data such as social media and Internet searches, these can be obtained either at an individual level through consenting participants or at a more aggregate level with tools such as Google Analytics and Trends. With the increasing availability of various sources of open data, archival data can easily be used by social and personality psychologists to either complement existing methods to ensure external validity and generalizability or to test more complex questions that would otherwise require years of primary data collection.

Big Data

With every passing year, technology becomes increasingly robust and adept at collecting massive amounts of data on an endless variety of human behavior. For scientists conducting research in social and personality psychology, these recent technological advances have led to the development of new tools that change how researchers can collect and analyze data. "Big data" are becoming a common term that refers not only to very large sets of data but also to the tools and techniques that are used to analyze it.

The three defining properties of big data include velocity (speed of data processing/collection), volume (amount of data), and variety (types of data). There is no definitive cutoff for each of these criteria (e.g., 10,000 data points do not necessarily make data "big"), however, a few examples include electronic medical health records, data generated by radio-frequency identification tags, and global positioning system data. Big data also can be collected through social media sites and search engines and often can include real-time data (Dufau et al., 2011; Manovich, 2011). It is important to distinguish big data from its aforementioned relative, archival data. An easy way to distinguish them is to recognize that while high volume, variety, and velocity archival data can be a form of big data, not all archival data are big data and not all big data are archival data. Archival data are merely any data that are not primarily collected by the researchers for the specific purpose of a study, while big data pertain to the scope and purpose of the data set.

Big data diminish the impact and intensity of the research dilemmas caused by common trade-offs between generality, control, and realism (Runkel & McGrath, 1972). Specifically, by utilizing big data, it is possible to create or isolate multiple conditions (thereby establishing control), collect data in natural

settings (providing high fidelity), and consolidate data from a large and diverse sample (generality). Here, we discuss two examples of emerging big data technologies—microexpression tracking software and wearable sensors—that offer unique opportunities for social and personality psychologists.

Microexpression trackers automatically and continuously code brief, unconscious facial expressions that indicate specific emotional reactions to stimuli (Ekman, 2009; Shreve, Godavarthy, Goldgof, & Sarkar, 2011). That is, video streams or recordings of facial expressions serve as the input to automated, simultaneous coding of fundamental emotions (Shreve et al., 2011). From a methodological perspective, this tool offers an objective and unobtrusive indicator of universally experienced emotions that may otherwise be undetected or misunderstood. From a practical perspective, automating the processes of microexpression detection and recognition lessens the burden on researchers and allows for the collection and analysis of continuous data, thereby deepening our understanding of the role of emotions across research questions in social psychology.

This technology has many applications, for example, it can be used to study criminality (Ghaziasgar, De La Cruz, Bagula, & Connan, 2016) as well as more generally to study unconscious affective responses to various stimuli (Barsade, Ramarajan, & Westen, 2009). Researchers appear to already be thinking about how this technology relates not only to psychological assessment but also to improving national security in our airports and other transportation hubs (Owayjan, Kashour, Al Haddad, Fadel, & Al Souki, 2012). Future research should leverage these developments in detecting our unconscious emotions when looking at momentary affective responses.

An emerging and powerful research tool in social and personality psychology is that of wearable technology (e.g., Fitbit or Apple Watch), which allows for the noninvasive collection of physiological and interpersonal interaction data on a massive scale. The origins of this technology can be traced to the field of psychophysiology, which has involved studying biological processes to understand behavior since the early 19th century (Lachman, 1965). Psychologists have asked participants to wear heart rate monitors to determine anxiety and arousal (Fowles, 1980) and tested saliva to determine cortisol (stress) levels (Kirschbaum, Prike, & Hellhammer, 1993). In the past, researchers have grappled with how to obtain biological measures (e.g., heart rate) in a real-time, noninvasive fashion (Mulder, 1992). Today this problem is all but solved: Many people are voluntarily wearing devices that track their physiological data, rendering the method noninvasive. Researchers have already harnessed the advantages of wearable technology in medicine and biomedical engineering (Bonato, 2003; Park & Jayaraman, 2003; Patel, Park, Bonato, Chan, & Rodgers, 2012; Vallurupalli, Paydak, Agarwal, Agrawal, & Assad-Kottner, 2013); however, social and personality psychologists have yet to leverage this technology in their research.

One potentially useful personal sensor is the sociometric badge. This specific wearable sensor allows for the analysis

of speaking, conversations, and body language in an interpersonal setting. Whereas the previously described sensors gather primarily biological data (e.g., sleep data and heart rate), the sociometric badge gathers information that is inherently interpersonal. The sociometric badge collects momentary-level data during interpersonal interactions or meetings. For example, an hour-long meeting with six individuals would have 3,600 s worth of data for numerous variables (e.g., length of time speaking; direction faced when speaking) for each of the six people—a massive quantity of data. This level of granularity and sensitivity is especially useful when studying intergroup processes, as it allows for new kind of team- and group-level analyses that are both more organic and reliable than observation or self-report. Still in its infancy regarding research application, the sociometric badge has been used in understanding consumer behavior patterns and assessing the effectiveness of marketing (T. J. Kim, Chu, Brdiczka, & Begole, 2009). By utilizing the badge in real time during meetings, the badge can enhance collaboration (T. Kim, Chang, Holland, & Pentland, 2008). Additionally, the data gleaned from sociometric badges have been shown to predict important outcomes such as employee reported satisfaction, productivity, and the objective quality of group interaction (Olguín Olguín, 2007).

The potential pitfalls of these new tools are the inherent risks in their technological sophistication (e.g., accuracy) and data quantity strengths (e.g., slow processing times; see Table 1 and Chaffin et al., 2017); researchers should be aware that technologies are imperfect and that massive amounts of data come with massive challenges. Indeed, Chaffin and colleagues (2017) took a systematic approach to examining the validity of one wearable sensor and concluded that excitement about its potential should be tempered with caution about systematic measurement errors. Concerns regarding the ethics and atheoretical nature of some forms of big data (see Tonidandel, King, & Cortina, 2015) should also be considered carefully.

(Good Old) Observations

Finally, we describe one of the earliest methodologies used in social science that of the observational method. In this method, there are often very few, if any, manipulations, and the goal is simply to systematically record participants' behaviors. Observational methods can be thought of in at least two broad categories. The first involve *participant observations*, in which the researcher asks the participants to recall; draw observations; or reflect on memories, events, and/or situations. Such information can involve more qualitative data such as diary accounts, unstructured interviews, and responses to open-ended questionnaires. The second involves *seminaturalistic* or *naturalistic observations*, in which the researcher has little, if any, consent from or influence over the participant, and the participant often does not know he or she is being watched. Gaining this lack of awareness from the participant achieves candid observations that are expressed in a natural setting.

Two of the most classic social psychological studies show the power and insights gained from observational methods. The

first, and one of the earliest, is the well-known study by Festinger, Riecken, and Schacter (1956) in which they studied an unidentified flying object cult (the “Seekers”), whose leader (Marian Keech) prophesied that the end of the world was coming on December 21, 1954. Because of his naturalistic observations, Festinger et al. (1956) were able to begin accounting for the conditions surrounding people’s continued beliefs about something in the face of disconfirming evidence. The second, and perhaps arguably one of the most well-known social psychological studies, is Zimbardo’s prison study (Haney, Banks, & Zimbardo, 1973) in which he and his team assigned individuals to the role of prison guard or prisoner and observed the result. Although Zimbardo did act as a warden, and he did assign people to play the role of prison guard or prisoner, many consider this a seminaturalistic observational study. The study monitored people over a series of days until it was called off—and it represents a compelling psychological demonstration showing that people behave in certain ways arguably due to their situation and not their disposition (American Psychological Association, 2004).

The choice of observational study participants ideally represents a randomized sample of the general population or general population of interest. For this reason, large, diverse (and naive) samples are ideal. In Zimbardo’s (1973) seminaturalistic prison study, the researchers advertised participation to the entire Stanford undergraduate population, used personality tests to screen out individuals with confounding traits, and then randomly assigned students into observational roles. Participants found via traditional, naturalistic observational methods are typically not required to provide their informed consent. In such cases, the onus is still on researchers to ensure that the locations being examined are truly public, safe, and do not interfere with the privacy rights of the participants (e.g., observing behaviors in restrooms).

The biggest strength of observational research is its inherent ecological validity; in no other kind of research can data be so clearly real (see Table 1). One can see this vividly through the knowledge that Ainsworth, Blehar, Waters, and Wall (1978) gained about attachment theory through watching toddlers interact with their mothers and then strangers in the *stranger paradigm* and, further, in the knowledge that Gottman and his colleagues gained about marital strength and dissolution from observing authentic couples engaging in natural behaviors (Gottman & Levenson, 1992; Levenson & Gottman, 1983). Of course, the value of observational methodology must be considered in light of some limitations that may explain why, despite such early beginnings as a very powerful methodological tool, observational methods in social psychology remain an underutilized tool.

Such studies can be complex, time intensive, and leave the researcher wanting more control. Additional ethical issues arise with the sampling of participants and the sampling of settings in observational studies. For instance, Hebl, Foster, Mannix, and Dovidio (2002) evaluated interpersonal bias against homosexual (vs. Texans) job applicants (i.e., confederates) in a natural work setting by having research assistants record their

interactions through discrete recording devices. This created an ethical dilemma true for many observational studies, as participants are often unaware that they are being studied. Moreover, it became the responsibility of Hebl et al. (2002) to decide the ethics of the sampling setting—whether the location was truly public or if it infringed on privacy rights. Another issue dealt with by these researchers (and countless others) was the sampling of behaviors and determining whether their sample of behavior is true for the full range of behaviors rather than unique to the isolated period of time in which they were observed.

Finally, another significant drawback of behavioral observation studies is that they can become very time-consuming. These studies often involve coding schemes that are highly complex and require researchers to agree on accurate definitions for the types of behaviors observed. Researchers must ensure that research assistants have an accurate understanding of coding schemes through intensive training. The duration of coding data may take months, or even years, to complete depending on the complexity of the study (see Hebl et al., 2002; Zimbardo, 1973). With good reason, social and personality researchers are deeply concerned about causal interpretations and such conclusions may be hard to come by in observational research in which these complications arise.

Designing observational studies typically begins with a decision about the level of control that will be asserted; researchers must determine the extent to which the context will be contrived. Another critical issue concerns the extent to which the study and researchers are known, versus hidden, to the participants (see Hong & Duff, 2002). Additionally, the safety of the researchers conducting the observations should be fully considered (see Adang, 2016). Upon its collection, observational data must be organized in a meaningful and interpretable manner through processes that vary in structure. Each of these decision steps influences the nature of the inferences that can be drawn. A highly controlled observation lacks realism but offers greater confidence in conclusions; an unobtrusive observation may be most genuine but also can involve ambiguity and deception, and the degree of analytic structure directly affects the breadth of conclusions that can be drawn. Resolving each of these decisions in line with the specific goals of a particular study can help to ensure that the observations ultimately describe the phenomenon of interest.

Conclusion

The call of this special issue was to focus on new methodological developments for social and personality psychology researchers. We believe previous findings have been limited by the types of methods that researchers have historically relied upon (e.g., the college laboratory setting) as well as by the newer methods that they are currently relying upon (e.g., mTurk and other survey platforms) to do their research. We contend that it is not necessarily new methodological developments that are needed for our field. Rather, we hope to have inspired researchers who rely more upon some of the oldest and

most underutilized methodologies that we have. In particular, we hope people gain interest, skills, and knowledge of the benefits gained by moving beyond the standard college laboratory and the SurveyMonkey or Qualtrics data survey collection platforms and into the real world. The strong allure of continuing to rely upon the stream of available college students or paid participants who “must” show up can lead us to become lazy in creating diversity in our data collection. However, it is up to us, as researchers, to ensure that we maximize the quality of data when testing our hypotheses. If we fail in this goal, we may not be getting valid answers to our questions of interest, especially whether we are tapping into populations that do not really care about providing good data in the first place. We encourage people to learn more about, consider adopting, and look for the ways that their research findings and generalizability may be expanded by using some of the most age-old methods (e.g., observations and field studies) as well as new methods (e.g., big data) to move our field forward. We recognize that the methods that we have highlighted in this article also have their own sets of limitations; however, we believe a greater reliance upon them as well as triangulation of data collection strategies ensure the field of social and personality psychology stays relevant and legitimate.

Authors' Note

All of these authors contributed equally to this article and so authorship is listed by alphabetical order.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Adang, O. M. (2016). A method for direct systematic observation of collective violence and public order policing. *Sociological Methods & Research*. doi:10.1177/0049124116661578
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.
- American Psychological Association. (2004). *Demonstrating the power of social situations via a simulated prison experiment*. *Psychological science: Research in action*. Retrieved from <http://www.apa.org/research/action/prison.aspx>
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63, 602.
- Barsade, S. G., Ramarajan, L., & Westen, D. (2009). Implicit affect in organizations. *Research in Organizational Behavior*, 29, 135–162.
- Bonato, P. (2003). Wearable sensors/systems and their impact on biomedical engineering. *IEEE Engineering in Medicine and Biology Magazine*, 22, 18–20.
- Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods*, 20, 3–31. doi: 10.1177/1094428115617004
- Crosnoe, R., & Elder, G. H., Jr. (2002). Successful adaptation in the later years: A life course approach to aging. *Social Psychology Quarterly*, 65, 309–328.
- Dholakia, U. (2017). Just how many Amazon mTurk survey-takers are there? *Psychology Today*. Retrieved January 30, 2017, from <https://www.psychologytoday.com/blog/the-science-behind-behavior/201701/just-how-many-amazon-mturk-survey-takers-are-there>
- Druckman, J. N., & Kam, C. D. (2009). Students as experimental participants: A defense of the ‘narrow data base.’ doi:10.2139/ssrn.1498843
- Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., . . . Ktori, M. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLoS One*, 6, e24974.
- Ekmann, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (Rev. ed.). New York, NY: W.W. Norton & Company.
- Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world*. Minneapolis: University of Minnesota Press.
- Fowles, D. C. (1980). The three arousal model: Implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology*, 17, 87–104.
- Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, 67, 92–96.
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57, 321–326. Retrieved from https://aom.org/uploadedFiles/Publications/AMJ/Apr_2014_FTE.pdf
- Ghaziasgar, M., De La Cruz, N., Bagula, A. B., & Connan, J. (2016). Visual data mining: A great opportunity for criminal investigation. In O. E. Isafiade & A. B. Bagula (Eds.), *Data mining trends and applications in criminal science and investigations* (pp. 112–141). Hershey, PA: IGI Global.
- Google. (2016). *What is the EU? What is Brexit?* Retrieved from <https://www.google.com/trends/explore?date=now%201-d&geo=GB&q=what%20is%20the%20eu,what%20is%20brexit&hl=en-US>
- Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology*, 63, 221.
- Haney, C., Banks, W. C., & Zimbardo, P. G. (1973). A study of prisoners and guards in a simulated prison. *Naval Research Reviews*, 9, 1–17.
- Hebl, M. R., Foster, J. B., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin*, 28, 815–825.
- Hebl, M. R., King, E. B., Glick, P., Singletary, S. L., & Kazama, S. (2007). Hostile and benevolent reactions toward pregnant women:

- Complementary interpersonal punishments and rewards that maintain traditional roles. *Journal of Applied Psychology*, 92, 1499.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83.
- Hernandez, I., Newman, D. A., & Jeon, G. (2015). Twitter analysis: Methods for data management and a word count dictionary to measure city-level job satisfaction. In S. Tonidandel, E. King, & J. Cortina (Eds.), *Big data at work: The data science revolution and organizational psychology* (pp. 64–114). New York, NY: Routledge.
- Hong, L. K., & Duff, R. W. (2002). Modulated participant-observation: Managing the dilemma of distance in field research. *Field Methods*, 14, 190–196.
- Kim, T., Chang, A., Holland, L., & Pentland, A. S. (2008, November). Meeting mediator: Enhancing group collaboration using sociometric feedback. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 457–466). Chicago, IL: ACM.
- Kim, T. J., Chu, M., Brdiczka, O., & Begole, J. (2009, April). Predicting shoppers' interest from social interactions using sociometric sensors. In *Extended abstracts of the 2009 ACM conference on computer supported cooperative work* (pp. 4513–4518). Chicago, IL: ACM.
- King, E. B., & Ahmad, A. S. (2010). An experimental field study of interpersonal discrimination toward Muslim job applicants. *Personnel Psychology*, 63, 881–906.
- King, E. B., Dawson, J. F., West, M. A., Gilrane, V. L., Peddie, C. I., & Bastin, L. (2011). Why organizational and community diversity matter: Representativeness and the emergence of incivility and organizational performance. *Academy of Management Journal*, 54, 1103–1118.
- King, E. B., Hebl, M. R., Botsford Morgan, W., & Ahmad, A. (2013). Experimental field research on sensitive organizational topics. *Organizational Research Methods*, 16, 501–521.
- King, E. B., Shapiro, J. R., Hebl, M. R., Singletary, S. L., & Turner, S. (2006). The stigma of obesity in customer service: A mechanism for remediation and bottom-line consequences of interpersonal discrimination. *Journal of Applied Psychology*, 91, 579.
- Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28, 76–81.
- Lachman, S. J. (1965). *History and methods of physiological psychology: A brief overview*. Detroit, MI: Hamilton Press.
- Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45, 587.
- Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *American Journal of Sociology*, 44, 868–896.
- Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. *Journal of Experimental Social Psychology*, 66, 100–106.
- Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the Digital Humanities*, 2, 460–475.
- Martin, L. R., Friedman, H. S., & Schwartz, J. E. (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology*, 26, 428.
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9, 147.
- Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34, 205–236.
- Olguín Olguín, D. (2007). *Sociometric badges: Wearable technology for measuring human behavior* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge.
- Owayjan, M., Kashour, A., Al Haddad, N., Fadel, M., & Al Souki, G. (2012, December). The design and development of a lie detection system using facial micro-expressions. In *2012 2nd international conference on advances in computational tools for engineering applications (ACTEA)* (pp. 33–38). IEEE.
- Paluck, E. L. (2011). Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, 47, 350–358.
- Paluck, E. L. (2016). How to overcome prejudice. *Science*, 352, 147–147.
- Paluck, E. L., & Cialdini, R. B. (2014). Field research methods. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 81–98). Cambridge, U.K.: Cambridge University Press.
- Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, 103, 899.
- Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113, 566–571.
- Park, S., & Jayaraman, S. (2003). Enhancing the quality of life through wearable technology. *IEEE Engineering in Medicine and Biology Magazine*, 22, 41–48.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530.
- Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 9, 1.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28, 450–461.
- Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method*. New York, NY: Holt, Rinehart and Winston.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.
- Shreve, M., Godavarthy, S., Goldgof, D., & Sarkar, S. (2011, March). Macro-and micro-expression spotting in long videos using

- spatio-temporal strain. In *2011 IEEE international conference on automatic face & gesture recognition and workshops (FG 2011)* (pp. 51–56). IEEE.
- Shultz, K. S., Hoffman, C. C., & Reiter-Palmon, R. (2005). Using archival data for I-O research: Advantages, pitfalls, sources, and examples. *The Industrial-Organizational Psychology, 42*, 31–37.
- Tonidandel, S., King, E., & Cortina, J. (2015). *Big data at work: The data science revolution and organizational psychology*. New York, NY: Routledge.
- Tucker, J. S., Friedman, H. S., Schwartz, J. E., Criqui, M. H., Tomlinson-Keasey, C., Wingard, D. L., & Martin, L. R. (1997). Parental divorce: Effects on individual behavior and longevity. *Journal of Personality and Social Psychology, 73*, 381.
- Vallurupalli, S., Paydak, H., Agarwal, S. K., Agrawal, M., & Assad-Kottner, C. (2013). Wearable technology to improve education and patient outcomes in a cardiology fellowship program—a feasibility study. *Health and Technology, 3*, 267–270.
- Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*, 493–504.
- Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition, 2*, 243–256.

Author Biographies

Ho Kwan Cheung is a graduate student at George Mason University.

Mikki Hebl is the Martha and Henry Malcolm Lovett Professor of psychology at Rice University. She is the corresponding author on this article and can be reached at hebl@rice.edu.

Eden B. King is an associate professor at George Mason University.

Hannah Markell is a graduate student at George Mason University.

Carlos Moreno is a graduate student at Rice University.

Christy Nittrouer is a graduate student at Rice University.

Handling Editor: Simine Vazire