



Into the wild: Field research can increase both replicability and real-world impact[☆]



Jon K. Maner

Kellogg School of Management, Northwestern University, United States

ARTICLE INFO

Article history:

Received 15 September 2014
Revised 28 September 2015
Accepted 29 September 2015
Available online 24 March 2016

Keywords:

Field research
Replication
Applied psychology
Research methods

ABSTRACT

Field research has the potential to substantially increase both the replicability and the impact of psychological science. Field methods sometimes are characterized by features – relatively high levels of participant diversity, relative lack of control over extraneous variables, greater focus on behavioral dependent variables, less room for researcher degrees of freedom, and lower likelihood of publication bias – that can increase the veracity and robustness of published research. Moreover, field studies can help extend psychological research in valuable ways to applied domains such as health, law, education, and business. Consequently, field studies, especially those that integrate an applied perspective, can provide information directly relevant for tackling important social problems. Incorporating field data into lines of basic research can increase not just the replicability, but also the relevance and impact of one's science.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

As researchers, most of us want to do work that is both replicable and impactful. Whether or not our work achieves these goals is undoubtedly influenced by the type of research designs we choose to adopt. When making decisions about what kinds of designs to pursue, researchers consider many different factors. Some methods, such as those involved in laboratory experiments, increase the rigor and control with which one can test hypotheses. Other methods, such as those used in qualitative research, allow researchers to delve deeply into rich narrative data sources provided by small numbers of participants. Still other methods, such as those involved in applied field research, allow researchers to evaluate questions of immediate relevance to solving important social problems. Indeed, every research design comes with its own unique set of strengths (and limitations).

This article describes a number of methodological features that, when incorporated into a line of research, may enhance both replicability and impact. Although these features are found in a range of different research settings (e.g., experimental, observational, laboratory, field), they may be especially common in field research. Consequently, this article presents arguments for integrating into programs of research a greater focus on collecting data in the field. Field research, particularly that which adopts an applied focus, provides researchers with valuable opportunities to deliver on social psychology's potential to make a real difference in facing some society's most recalcitrant problems including health disparities, climate change, ethnic prejudice, and economic

inequality (e.g., Klein, Shepperd, Suls, Rothman, & Croyle, 2015). When it incorporates features that increase the robustness of its findings – features that are the focus of this article – field research has the potential for dramatically increasing both the replicability and the impact of social psychological science.

2. Replicability and impact: complementary goals

Social psychological studies have been criticized in recent years for lacking replicability. Many factors have contributed to the current situation, including the use of questionable statistical practices, incentive structures in the publication system that reward positive results and “perfect” patterns of data, and the fact that some empirical findings may replicate only under certain circumstances (Giner-Sorolla, 2012; Ioannidis, 2005; John, Loewenstein, & Prelec, 2012; Maner, 2014; Simmons, Nelson, & Simonsohn, 2011; cf. Stroebe, 2016–in this issue).

A range of recent papers provide valuable suggestions aimed at evaluating and increasing replicability (e.g., Fabrigar & Wegener, 2016–in this issue; Sagarin, Ambler, & Lee, 2014; Sakaluk, Williams, & Biernat, 2014; Schaller, 2016–in this issue; cf. Finkel, Eastwick & Reis, 2015). Although producing research that is replicable is an important goal, however, it is not our only goal. One goal that at times has gotten lost in the chorus of voices involves the real-world impact and relevance of our science. Some would argue that an important function of social psychology is addressing real-world social problems such as climate change, public health, racial and ethnic prejudice, terrorism, and the growing divide between the haves and the have-nots. This article describes ways in which field research methods can increase the replicability of social psychological research, while also enhancing its real-world impact.

[☆] The author thanks Vlad Griskevicius, Sam Sommers, Loran Nordgren, and Mary Gerend for their helpful input.

E-mail address: jon.maner@kellogg.northwestern.edu.

The issues of replicability and impact are independent but also inherently hierarchical. A set of findings cannot be impactful if the findings are not replicable, but research certainly can be replicable, but not impactful. The relationship between replicability and real-world impact is analogous in some ways to the relationship between reliability and validity. Reliability (in the psychometric sense) comes first: a particular self-report measure, for example, cannot be valid if the items that comprise it do not hold together in a coherent way. But researchers usually are not satisfied with demonstrating a measure's reliability; they also need to show that it is valid and conceptually meaningful. The same goes for scientific impact. Demonstrating that a set of findings is replicable is not enough; those findings also need to advance the field, and one way to do that is to bear upon some substantive problem or issue in the world.

One concern is that some of the changes the field is generating to deal with issues of replicability may inadvertently reduce the potential for impact. For example, the widespread reliance on Mechanical Turk allows researchers to collect large samples quickly and easily and so helps address the important issue of statistical power. However, because such methods sometimes rely on having "professional" participants provide hypothetical responses to imagined scenarios, lines of research that depend heavily on such methods may be less impactful than those that include more immediate and direct measurement of actual behavior from naïve participants. The press for larger samples may also lead researchers to rely on self-report measures instead of measurement of behavior. One can often collect many self-report data relatively quickly, but measuring the way individuals actually behave in response to a particular situation can be more time-consuming. Nevertheless, self-report data, especially in the form of imagined responses to hypothetical scenarios, often only weakly predict behavior (e.g., Gollwitzer, Sheeran, Michalski, & Seifert, 2009; Sheeran, Abraham, & Orbell, 1999) and so may not be overly informative when behavior is the true outcome measure of interest. Some would argue that measuring actual behavior is essential for maintaining the health of our field (Baumeister, Vohs, & Funder, 2007). This is an important point: as a field, we should not focus on replicability to the detriment of impact.

Consistent with Cialdini's (1980) view of "full-cycle social psychology," programs of research might ideally include a combination of rigorous theory testing, controlled laboratory methods to identify underlying mechanisms, and applied field research to assess how psychological phenomena unfold in natural contexts. The integration of field methods provides valuable opportunities to tackle questions of great importance to society. This ultimately means getting out of the lab, directly measuring behaviors relevant to important real-world issues, and connecting with people in fields such as medicine, education, law, and business. Such an approach can also further the field's goals pertaining to replication: field research is particularly amenable to features that make it more replicable than studies relying on laboratory or online methods alone. Thus, collecting data in the field can achieve two goals simultaneously — it can enhance real-world impact and it can increase replicability.

3. Using field research to increase replicability

The following sections describe methodological features that may enhance the replicability of published research.¹ These features can be incorporated into many different types of research designs, but are especially common in field studies. Field studies can be operationally defined as observational or experimental studies that take place in settings such as schools, homes, the workplace, and health clinics — mundane settings in which people lead their lives naturally and that are not designed for the purposes of research. Field studies are valuable

in part because they afford opportunities to examine psychological and behavioral processes as they unfold in ordinary contexts (Cialdini, 1980). Nevertheless, field settings do not constitute a bounded category of research design. One might characterize research settings as existing on a continuum, with field settings such as the aforementioned at one end and studies occurring exclusively in the laboratory on the other. The remainder of the continuum reflects many other manner of research setting including archival studies, retrospective self-reporting of experiences from the field, and data conducted via social media such as Twitter or Facebook.

Studies toward the field end of the continuum are sometimes characterized by features that, all else being equal, may enhance their replicability. Field methods often: (1) use participant samples that are diverse; (2) exert less control over extraneous variables; (3) focus on the direct measurement of behavioral dependent variables; (4) allow for fewer researcher degrees-of-freedom that can hinder replicability and (5) are less likely to produce research literatures that suffer from publication bias (see Table 1).

Not all field studies are characterized by these features, because studies inevitably vary in their specific methodological details. Moreover, such features may also be present in other types of designs (e.g., laboratory research), particularly if researchers exert effort to incorporate them into their work. Thus, field studies are not inherently more replicable than other types of designs; they are more replicable to the extent that they incorporate features such as these into their design. When these features are incorporated into a line of field research, however, the result can be a highly replicable and impactful piece of science.

3.1. Participant diversity

Studies that rely on field methods can involve participant samples that are more diverse than those used in the lab. Laboratory studies often make use of undergraduate participant samples that are relatively homogenous. At the very least, those samples consist of participants who are all students at the same university, most of whom are of similar age and who have elected to sign up for an undergraduate psychology course. Similarly, many studies conducted via online data collection platforms (e.g., MTurk) consist of participant samples that have

Table 1

Field studies sometimes provide a number of advantages that increase the replicability of psychological science.

Replicability issue	Feature of field research	Advantages
Participant diversity	Field studies often employ samples that are highly diverse	Findings derived from diverse samples are more robust and may be more likely to replicate across other samples than those derived from homogenous samples
Presence of extraneous variables	Field studies usually entail less control over extraneous sources of variance	Effects demonstrated under uncontrolled circumstances should be more robust to contextual factors than those in which such factors are held constant
Measurement of behavior	Field studies often focus on directly measuring behavioral dependent variables	Behavioral DVs in field studies are relatively robust to variables that might moderate the intention-behavior gap, and thus may be more replicable (and impactful)
Researcher dfs	Field studies sometimes focus on fewer DVs; entail less daily control over data collection	Field studies may leave relatively less room for researcher dfs
Publication bias	Field studies tend to be high-investment, less likely to be relegated to the file drawer	Field studies may be less likely than other types of research to suffer from publication bias

¹ When I refer to replicability I am referring to forms of both direct and conceptual replication. Although there are important differences between the two, most of the issues discussed pertain to both forms of replication.

participated in large numbers of online studies and so are homogenous at least with respect to having sought ready access to an online data collection platform and being very well-acquainted with the methods and hypotheses used by behavioral researchers (Rand et al., 2014; see also Berinsky, Margolis, & Sances, 2016—in this issue and Curran, both 2016—in this issue).

Studies conducted in the field, in contrast, often rely on samples that are more naïve and that display greater demographic diversity and individual difference variability. For example, Knowler et al. (2002) conducted a field trial in which they assessed effects of lifestyle changes (diet, exercise) on type 2 diabetes onset. Their study was conducted in 27 clinical centers around the country and, in addition to amassing an impressively large sample ($n = 3234$), 45% of their sample included members of underrepresented minority groups. By virtue of its setting, the study therefore included a sample of participants that was considerably more diverse than one might obtain using other methods. Although not all studies are able to include such an impressively large and diverse group of participants, participant diversity is an important goal in part because findings from studies using diverse samples may be relatively robust to individual differences that covary with ethnic and racial group status.

Indeed, a study's level of participant diversity has direct implications for its replicability. If some effect is observed in a diverse set of participants, that effect should, on average, replicate more often than one that involves a homogenous set of participants. Most psychological phenomena involve processes that vary to some degree across participants and that variance is tied to stable attributes of the person. Many processes are moderated by individual difference variables such as personality traits and demographic characteristics. One reason some studies may fail to replicate is that there are “hidden” moderating variables; perhaps the effect being investigated replicates only in people who are high or low on a particular trait left unmeasured in the original study (see Fay & Maner, 2012, for a discussion). A study that relies on a diverse sample provides opportunities to assess whether the findings are robust to such moderating variables. For example, if an effect is demonstrated in a sample of participants who vary considerably in their race or socioeconomic status, any moderating effect of race or SES can be identified and taken into account in future research. If no moderating effects of race or SES are found, the effect is likely to be robust to variability in such factors. Either way, the use of a diverse sample arms researchers with information that can aid in attempts to replicate and extend the research. Contrast this with a study that relies on a sample that is homogenous with respect to race or SES. It would be difficult to know whether effects demonstrated in such a study are robust to variability in those factors. Consequently, if that study were re-created in a different or more diverse sample, replication is less certain because the new sample might include greater variability on individual differences that moderate the effect. Consequently, all else being equal, effects demonstrated in diverse samples should be more replicable than those from homogenous samples. Thus, the use of diverse participant samples – a hallmark of many field studies – can increase a study's likelihood of being replicated.

Many replications are designed to recreate the original procedures and sample as closely as possible. If a lab performs a direct replication in the same population from which the original study was drawn, this tells us something about whether the findings replicate in people and circumstances that closely mirror the original study. What also matters, however, is whether a pattern of findings replicates more generally – in people and circumstances that reflect normal variability in the world and diverge from the original study (Finkel et al., 2015; Klein et al., 2015). Such replications provide greater confidence in the overall robustness of the findings. Although studies relying on homogeneous samples of laboratory or online participants might be highly replicable when conducted again in a similar homogeneous sample of laboratory or online participants, this is not the key criterion (or at least not the only criterion) on which we should judge replicability (Westfall, Judd,

& Kenny, 2015; see also Brandt et al., 2014; Stroebe & Strack, 2014). Just as important is whether studies replicate in samples that include participants who reflect the larger and more diverse population of people for whom the psychological phenomenon is relevant. Indeed, issues of replication are inextricably linked to issues of generalizability and robustness. Because field studies typically rely on participant samples that are relatively diverse, they are likely to be relatively robust to variability in sample characteristics. Moreover, with greater participant diversity also comes greater opportunities to assess potential moderating effects of individual differences. With these advantages in mind, it would behoove researchers working in the field to increase the diversity of their participant samples whenever possible and, ideally, recruit representative samples. This would potentially provide the greatest increase in replicability across samples.

3.2. The role of extraneous variables

Participant characteristics are not the only source of variability relevant to the issue of replicability. A study's replicability is influenced also by the presence of any number of other situational and contextual variables. One strength of laboratory research is that it allows researchers to control extraneous variables, to eliminate unnecessary and undesirable sources of error variance, and to narrow the situation as much as possible so that researchers can hone in on those manipulations and measures that are most central to one's theory. This approach provides rigorous and highly controlled tests of one's hypotheses and thus is a critical piece of theory testing in behavioral science (Mook, 1983).

Recent calls regarding replicability may further encourage researchers to reduce the presence of situational variability. Doing so can increase effect sizes and thus reduce the burden of collecting very large participant samples. Reducing unnecessary variability within a study also produces a streamlined procedure and so can be more easily and accurately re-created by other researchers who seek to replicate one's work.

However, the highly controlled nature of laboratory research also comes with a cost, one that may inadvertently reduce the replicability of one's work. Because laboratory research methods are designed in part to rule out extraneous variables, they reduce the likelihood with which findings will replicate in any context in which those variables are allowed to vary. That is, rigorous laboratory studies control many of the factors that may be more variable in other labs or in naturalistic contexts. When allowed to vary, those factors can increase the relative amount of error variance, which could reduce effect sizes and obscure the effect being investigated.

Moreover, such factors could moderate the effect being examined. For example, a research team might routinely employ practices designed to evoke a neutral mood in participants (e.g., having research assistants behave in a dispassionate fashion). Many psychological phenomena are moderated by mood (e.g., judgment accuracy declines when people are in a positive mood; Alloy & Abramson, 1979), so if that research team produces a finding that is then recreated in a different lab in which research assistants are allowed to interact more freely with participants, the pleasantness of those interactions could moderate the effect and undermine replication. In sum, controlled laboratory findings may fail to replicate when studies are moved from tightly controlled contexts into more variable contexts.

Indeed, even very subtle factors such as the ambient temperature or lighting in the room can play a surprising role in shaping behavior (Schaller, Park, & Mueller, 2003). Such factors can add error variance or serve as covert moderating variables that, when allowed to vary, hinder replicability. Myriad factors that were controlled in an original study, particularly those that are unmeasured, may obscure findings when attempting to replicate those findings in another laboratory or in the field. Thus, effects that have been demonstrated under highly controlled circumstances should, on average, be less replicable than those that have been demonstrated by researchers exercising less

control over extraneous variables. Because it occurs in relatively natural contexts, field research involves less control over extraneous factors and, therefore, results of field studies should be relatively more replicable because they should be relatively more robust to extraneous variables.

For example, one field study conducted in Rwanda assessed whether exposure to mass media can reduce prejudice (Paluck, 2009). The study manipulated whether people were exposed to a daytime radio show featuring conciliatory themes. Aside from assuring that control participants, who were assigned to listen to a radio show about health, remained untreated, little effort was devoted to controlling extraneous factors. Data were collected over the course of one year, and during that time people went about their lives in a natural fashion. Thus, countless variables extraneous to the study were permitted to vary. Consequently, the findings, which revealed significant evidence for the efficacy of the radio show, provide evidence of mass media effects that are likely to replicate in a broader range of circumstances than those produced in a highly controlled context. The effect of the radio show was observed despite the presence of many extraneous factors, potential moderating variables, and sources of error variance. Because effects demonstrated in the field are relatively robust to such sources of variability, they are likely to generalize in “less-than-ideal” circumstances and across contexts in which such factors are allowed to vary.

Field studies may also provide estimates of effect size that are particularly useful, in the sense that they can be more readily generalized across a range of situations. Laboratory effect sizes say very little about the magnitude with which effects are likely to unfold outside the laboratory. For example, effect sizes obtained from laboratory experiments rely heavily on the strength of the particular manipulations selected. Field studies, in contrast, provide estimates of effect sizes that are likely to reflect the way phenomena unfold under natural circumstances.

Field studies provide powerful opportunities to demonstrate the real-world relevance and importance of the phenomenon under investigation. When coupled with rigorous experimental data, such studies provide a vivid portrait of how such phenomena unfold in naturalistic settings. Integrating field data into a program of research, for example by replicating experimental studies in a field context, not only increases the impact of social psychological science, it also provides compelling evidence for robustness and generalizability—issues of central importance to a phenomenon's replicability. Because field settings often involve variability in both participant characteristics and situational variables, they provide useful opportunities to develop theories that specify how particular manipulations interact with individual differences and contextual moderating variables.

3.3. Measuring behavior

A key issue identified as contributing to false positive findings is the prevalence of under-powered studies (Francis, 2012). Indeed, because a lack of statistical power can cause spurious results and inflated measures of effect size, it has been held up as possibly the most critical linchpin of the replicability crisis. Large, highly powered studies mitigate concerns readers might have about results reflecting questionable research practices.

Indeed, the recommendation arising from discussions of statistical power seems clear: run studies with larger samples. But this recommendation comes with a potential downside. Collecting a large amount of behavioral data can be very costly, entailing a substantial expenditure of resources including time, money, personnel, and space. This is the case especially for difficult to run studies involving real social interactions or the direct assessment of interpersonal behavior. One reaction to the call for higher-powered studies might be to avoid such studies, precisely because they are so costly when sample sizes need to be high. Instead many researchers might turn to self-report studies or simpler online data collection methods, which allow for the collection of large samples with relatively little effort or cost.

To be clear, self-report measures and online samples reflect valuable methodological tools. However, when research relies too heavily on such tools, it can undermine one of the basic goals of our science — to produce findings that have real-world impact. Although reliance on such methods may enable researchers to collect large samples with relatively few resources, and so satisfies the call for greater statistical power, it may inadvertently produce findings that lack impact. Such methods can contribute to an already existing trend in personality and social psychology away from measuring actual behavior and instead toward relying exclusively on self-report (Baumeister et al., 2007). Psychological processes can be studied effectively using a variety of different methods. Nevertheless, exclusive reliance on introspective methods and self-report variables has been an ongoing weakness of our field and it has caused research to become disconnected from many of the outcomes that really matter. Studies that measure actual behavior are critical for the health and relevance of our work.²

Because field research often focuses on applied research questions, such research is often designed to assess behavioral outcomes (Klein et al., 2015). Measures of behavior (e.g., getting a mammogram) may be harder to obtain than self-report measures (e.g., saying you would get a mammogram if it were offered), but manuscripts that include behavioral measures are likely to be more replicable, not to mention more impactful. Most researchers recognize the existence of the intention-behavior gap (e.g., Gollwitzer et al., 2009). There are intervening steps that translate intentions into behavior, and those steps can be interrupted by any number of moderating variables. By measuring actual behavior one provides findings that may be relatively robust to the moderating effects of those variables. For example, although asking people whether they would behave aggressively in response to a hypothetical situation may be useful for theory testing, it is difficult to know whether such responses would translate into actual aggressive behavior. Any number of factors present in the situation could prevent aggressive intentions from eliciting aggressive behavior (e.g., the salience of social norms against aggression). Demonstrating an effect on actual aggressive behavior shows that the effect in question is robust enough to “cut through” those factors.

If all intervening factors need to be at optimal levels to demonstrate an effect on behavior, then an effect demonstrated under such optimal circumstances might weaken when moved to contexts in which those intervening factors are allowed to vary. However, as discussed earlier, field settings tend to be more variable, not less, and so effects on behavior that are demonstrated in the field should be relatively robust to variance in intervening factors. Thus, if a field study were to demonstrate effects on a behavioral measure of aggression, one might surmise that the effect is robust enough to have overcome intervening psychological variables that moderate the link between aggressive intentions and actual aggression.

My intention is not to eschew self-report variables or other non-behavioral measures. Such measures are a valuable part of psychological science and measuring cognitive processes can be valuable in its own right. My intention is to caution researchers against allowing the call for greater statistical power to discourage them from using high-impact behavioral methods. Applied work from the field often deals directly with behavioral outcomes (e.g., academic achievement, health behavior, legal decisions), and researchers, especially those working in applied settings, should be encouraged to integrate direct assessments of behavior. Such assessments can increase not just the replicability, but also the impact of one's work.

3.4. Reducing researcher degrees of freedom and publication bias

Another set of factors that undermines replicability involves researcher degrees of freedom and publication bias. In responding to

² The question of how to define behavior is itself debatable. For example, one might include reaction time measures or retrospective reports of behavior as “real” behavior, while eschewing measurements of behavioral intentions or responses to hypothetical scenarios.

implicit incentives for reporting “perfect” data, researchers sometimes attempt to iron out all the wrinkles in their findings (Giner-Sorolla, 2012; Maner, 2014). This includes being selective in the reporting of dependent variables, exploiting the inclusion of post-hoc covariates, starting and stopping data collection based on frequent observation of the data, and failing to report entire studies in which patterns of data do not adhere to one’s hypotheses (John et al., 2012; Simmons et al., 2011).

Relative to other types of methods, field research methods may be less likely to suffer from exploited researcher degrees of freedom and publication bias. This is due to the fact that, relative to other types of methods, field research methods often focus on smaller numbers of key measures; they sometimes entail less control over the day to day supervision of studies; and they usually involve higher overall levels of investment than do laboratory or online studies.

In field studies, one is often interested in a clear primary outcome measure. For example, in designing interventions that encourage conservation, Goldstein, Griskevicius, and Cialdini (2011) compared a number of different tactics, each designed to encourage hotel patrons to re-use their bathroom towels. In this study, the dependent variable was quite clear: did patrons re-use their towels? Those data were collected by room attendants and coded dichotomously (patrons did or did not re-use their towels). Field studies of weight loss generally focus on amount of weight lost as the dependent measure; studies of vaccination behavior focus on whether or not someone gets vaccinated; studies designed to increase academic achievement use clear measures of academic achievement such as GPA (Gerend & Shepherd, 2012; Klein et al., 2015; Stephens, Hamadi, & Destin, 2014). Because field studies such as these often focus on a clear dependent variable, there is relatively little room for researchers to exploit the possibility of multiple dependent variables in order to obtain significant results. In a lab or online study, a researcher might measure half-a-dozen or more possible dependent measures, and then selectively report only those that adhere closely to hypotheses. Indeed, recent data suggest that approximately two-thirds of researchers in social psychology engage in this practice (John et al., 2012). Secondary analysis of data from the field can of course be susceptible to the same issue, as researchers may mine datasets in search of noteworthy findings. Nevertheless, field studies that focus on only one or a small number of dependent variables may be relatively more replicable than studies that include many dependent variables.

Moreover, when field studies involve partnering with exogenous entities such as schools, hospitals, or businesses, they are entrenched within the normal day-to-day operations of those organizations. Due to their high level of embeddedness within a naturalistic context, it is far more difficult for such studies to be stopped and started based simply on a researcher’s observation of the data as they are collected. This practice has been identified as a key factor that can undermine a study’s replicability (Simmons et al., 2011). Thus, field studies may be relatively less susceptible to artificially stopping data collection in order to obtain a significant effect and, more broadly, may be less susceptible to forms of “p-hacking.”

Field studies may also be less likely than online or laboratory studies to suffer from publication bias — the tendency for researchers to publish positive results, while relegating to the file drawer studies that fail to support hypotheses or otherwise do not fit neatly into a packaged set of studies. Publication bias can hinder replication efforts, because published manuscripts fail to accurately reflect the larger set of research from which the published studies were drawn. Publication bias is problematic in part because it overestimates effect sizes and paints a portrait of phenomena as considerably cleaner and more robust than they really are (Francis, 2012).

Consider again Goldstein and colleagues’ hotel intervention study. Given the high expenditure of resources needed to conduct such a study, it seems unlikely that the study was accompanied by several other “failed” attempts at detecting an intervention effect. It seems

unlikely that a researcher who has invested tremendous time, energy, and money into conducting an uninformative field study would react by running another very similar study in the hope that the second study would conform better to hypotheses.

This can be contrasted with the publication of traditional laboratory experiments. Such studies, particularly those conducted at institutions with large participant pools, often can be conducted quickly, cheaply, and with relative ease. Even large sample studies relying on online data platforms such as MTurk can be conducted literally in one day, which presumably makes it easier to relegate such studies to the file drawer.

Thus, the ease and speed with which data can be collected may be linked to the likelihood of publication bias. Relative to research methods that afford data collection without a large expenditure of time and resources, field methods may be less likely to suffer from publication bias. Findings arising from field research thus should, on average, depict findings that provide more accurate measures of effect size and are more likely to be replicated.

Because field studies are often expensive and labor-intensive, researchers who undertake them are probably highly motivated to publish them. On one hand, it is therefore conceivable that researchers may be willing to engage in questionable research practices in order to convince reviewers and editors that their field data should be published. On the other hand, because field intervention studies often are so valuable to the field, they can sometimes be published even without evidence for statistically significant effects. Field trials often are publishable regardless of their outcome, which should reduce the tendency to engage in questionable research practices. For example, some intervention journals publish reports of “failed” interventions, thus mitigating the problem of publication bias. Readers would be well-advised to keep both possibilities in mind when considering the results of field research.

4. Using field research to enhance the real-world impact of psychological science

Field research may enhance not only the replicability of psychological science, but also its real-world impact. Social psychology is very well-equipped to tackle important social problems and to engage in applied translational research. From public health to climate change, the tools of social psychology are ideally suited to help practitioners and policy makers understand the root causes underlying social problems and design interventions aimed at promoting the welfare of people and their communities. To remain relevant, and to have real impact, our discipline needs to be more applied, and that means getting out into the field (Cialdini, 2009).

Lines of research that integrate basic science with applied field research provide valuable opportunities to test hypotheses in ways not afforded by basic science methods alone. Field studies allow researchers to test interactions between psychological processes and real-world contexts, which cannot always be adequately recreated in the lab (Cialdini, 2009; cf. Hüffmeier, Mazei, & Schultze, 2016—in this issue). Field studies, especially those that examine intervention effects longitudinally, also allow researchers to examine how psychological processes unfold over time (e.g., Stephens et al., 2014). Such lines of research are also broader in the scope of empirical evidence they provide and, as such, should be more robust than lines of research that rely on a more limited set of methods. Consider Fast and Chen (2009) who, in examining the relationship between power and aggression, reported a combination of controlled experiments in the laboratory along with data collected in the field using participants in the workplace. This allowed them to examine the effect of both manipulated power (in the lab) and actual power (a person’s role within a workplace organization) on aggression. Their approach showed not only that power can promote aggressive behavior, but also depicted ways in which this phenomenon unfolds in natural settings outside the lab.

Similar examples can be found in other areas of social psychology. For instance, from the literature on social class comes research demonstrating that, with interventions that use role models to communicate ways of overcoming class barriers, researchers can virtually eliminate the social class gap in academic achievement (Stephens et al., 2014). Stephens and colleagues exposed first generation students (as well as more traditional students) to role models who discussed how they had leveraged their lower SES background to promote social support networks that helped them succeed in school. Exposure to those role models (relative to role models who did not discuss the benefits of their lower SES background) led the first generation students to display better academic performance (as measured by the students' grade point average) over the academic year. This is the sort of field study that could not simply have been conducted in the lab, and it contains many of the features discussed elsewhere in this article: it included a diverse sample who varied on SES; it allowed extraneous factors in the situation to vary; it utilized a behavioral dependent variable; and it provides the sort of high-impact finding that can be used to address a critical social problem (academic underperformance by members of low SES groups).

Enos (2014) performed a randomized controlled field trial to assess the effects of intergroup conflict on attitudes toward racial outgroups. He assigned pairs of native Spanish-speaking confederates to wait on train platforms used by relatively homogenous groups of White participants living in Boston. The interracial contact was thus experienced incidentally and experienced over a relatively extended period of two months. Contrary to what might be expected based on some laboratory tests of the "contact hypothesis" (see Dovidio, Gaertner, & Kawakami, 2003), exposure to the Spanish-speaking confederates led White participants to display more negative and exclusionary attitudes toward immigrants. This study provides practical information about the causal effects of interracial contact in natural settings that could not be ascertained simply from conducted research in the lab.

From field research on self-regulation comes evidence suggesting that simple manipulations such as having people write down the time of day at which they will perform some health behavior can have enormous positive effects on public health (e.g., Milkman, Beshears, Choi, Laibson, & Madrian, 2011). From field studies on social cognition and memory comes research identifying variables that influence eyewitness testimony in legal contexts; this research provides means of enhancing the accuracy of eyewitness testimony (e.g., use of blind administrator feedback to prevent bias; presentation of photos sequentially rather than simultaneously; Bradfield, Wells, & Olson, 2002; Lindsay & Wells, 1985). This research has been tested, replicated, and implemented successfully in applied law enforcement settings nationwide, and has led to substantial legal reform. Across a vast range of research literatures, field studies produce impactful findings that have immediate relevance to solving important social problems faced by society.

Indeed, there is a long history of impressive field research in social psychology. Many seminal studies in social psychology – Sherif's Robbers Cave study (Sherif, Harvey, White, Hood, & Sherif, 1961), Aronson's jigsaw classroom research (Aronson, Blaney, Stephen, Sikes, & Snapp, 1978), and Rosenthal and Jacobson's (1968) work on self-fulfilling prophecies, just to name a few – were all conducted in the field. These studies have stood the test of time in part because they provided immediate solutions to key problems and also because they involved working with real people in natural settings.

One way to incorporate field methods into a program of research is to take findings from the lab and replicate and extend them in real-world settings (Cialdini, 1980; Mortensen & Cialdini, 2009). One might take evidence amassed from tightly controlled experimental work conducted in the lab and attempt to replicate that evidence in the field using a context and participant population for whom the phenomenon is particularly relevant. For example, Correll, Park, Judd, and Wittenbrink (2002) provided carefully controlled laboratory evidence for biased responding toward African-American targets in a first-person shooter paradigm. Plant and Peruche (2005) replicated and

extended that work in the field using samples of police officers – a sample for whom shooter biases are especially important – and showed that with training, officers were able to reduce biases toward shooting unarmed African-American targets.

Getting into the field and incorporating an applied focus also means interacting with other disciplines. Because it relies on theoretical frameworks that encompass a vast range of behavior, social psychology can interface effectively with many applied disciplines including law, medicine, business, and education. Many lines of social psychological research would be enriched by including teams of researchers that include people from applied fields. Such fields would benefit from social psychology's powerful ability to identify underlying psychological mechanisms. And social psychology would benefit from opportunities to replicate its work in ways not afforded by basic science approaches alone.

Researchers should consider the importance of working in the field when initially designing their research. Doing so could lead investigators to develop lines of research that look different from those generated when researchers focus exclusively on the basic science value of their work. For example, decision-making researchers can increase the impact of their investigation by incorporating dependent variables that reflect the phenomenon of interest as it occurs in applied contexts (e.g., direct assessment of health decision-making; Gerend & Shepherd, 2012). Researchers might also incorporate special participant samples that comprise groups for which the phenomenon under investigation is of particular importance. For example, in testing interventions aimed at increasing academic achievement, one might focus on groups that traditionally underperform, such as first-generation university students (e.g., Stephens et al., 2014). Although doing so may reduce variability in participant characteristics and thus might limit the extent to which findings from such a study would generalize to other populations (see the section above titled *Participant diversity*), such an approach could nevertheless increase the overall impact and relevance of the investigation. This reflects a case in which the increased impact of using a special population may offset the potential reduction in generalizability.

One benefit of incorporating field studies into lines of psychological science pertains to the ease with which the work can be disseminated to entities outside the field. Although controlled laboratory studies are a critically important part of our field, they sometimes lack ecological validity, and consequently the relevance of such studies can be difficult to communicate to non-social psychologists. Studies conducted in the field highlight the value and relevance of our work in a way that is more easily digestible by lay audiences. Adopting such an approach would facilitate interactions and dialog between social psychologists, practitioners, and policy-makers.

Indeed, psychological scientists could do more to interface with such entities, and working in the field affords greater ability to disseminate one's work beyond traditional academic outlets. Although recent efforts have focused on disseminating our work to the public via the media, the field ought to also communicate our work directly to policy makers and practitioners. Communicating with the media is a positive step, in part because it serves as a bridge to entities that might benefit from knowing more about our science. But our field could also do more to disseminate our science directly to the people in a position to deploy it (e.g., law-makers, educators, physicians). Collaborative partnerships with such entities could dramatically increase the impact of our science. Moreover, drawing upon such partnerships in planning our studies could enhance the design, implementation, and replicability of our research.

5. Conclusion

Most researchers want to do work that is valid and replicable. Most researchers also want to do work that matters and has real-world impact. Incorporating field research methods – particularly those with an applied component – into one's program of research has the potential

to achieve both of these goals. Studies that rely on field methods can enhance both the replicability and the impact of one's work.

To be sure, field studies often involve a larger expenditure of resources than do studies conducted on MTurk or in the lab. However, such studies are likely to be highly compelling, replicable, and impactful. The time and effort sometimes required to collect data in the field may prove to be a wise investment, both for the individual researcher and for the field. For reasons discussed in this article, field studies can provide valuable empirical and theoretical contributions, while also generating concrete solutions to important social problems.

This article is not at all meant to suggest that researchers should stop doing controlled lab research or collecting data online. Rather, the argument is that scholars would benefit from incorporating field research methods, if they are not doing so already. The best programs of research might ideally include some combination of rigorous theory testing (which often includes experimental methods to directly identify underlying mechanisms) and application (e.g., field studies, use of special samples, applied dependent variables). Furthermore, during the editorial process, reviewers and editors should evaluate lines of research as a function of the research's ability to contribute to applied science, as well as its ability to contribute to theory.

Field studies can enhance both the replicability and veracity of psychological science. Getting out into the wild and testing hypotheses in the field provides a relevant and finely tuned understanding of human behavior. And field research can increase the field's impact, providing information essential for solving important problems facing people and society.

References

- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and non-depressed students — Sadder but wiser. *Journal of Experimental Psychology: General*, *108*, 441–485.
- Aronson, E., Blaney, N., Stephen, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills: Sage.
- 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.
- Berinsky, A., Margolis, M., & Sances, M. (2016). Can we turn shirkers into workers? *Journal of Experimental Social Psychology*, *66*, 20–28 (in this issue).
- Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology*, *87*, 112–120.
- Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224.
- Cialdini, R. B. (1980). Full-cycle social psychology. In L. Bickman (Ed.), *Applied social psychology annual*. 1. (pp. 21–47). Beverly Hills, CA: Sage Publications.
- Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, *4*, 5–6.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *6*, 1314–1329.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19 (in this issue).
- Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup contact: The past, present, and future. *Group Processes and Intergroup Relations*, *6*, 5–20.
- Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 3699–3704.
- Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68–80 (in this issue).
- Fast, N. J., & Chen, S. (2009). When the boss feels inadequate power, incompetence, and aggression. *Psychological Science*, *20*, 1406–1413.
- Fay, A., & Maner, J. K. (2012). Warmth, spatial proximity, and social attachment: The embodied perception of a social metaphor. *Journal of Experimental Social Psychology*, *48*, 1369–1372.
- Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic: Considerations with the case of relationship science. *Journal of Personality and Social Psychology*, *108*, 275–297.
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin and Review*, *19*, 151–156.
- Gerend, M. A., & Shepherd, J. E. (2012). Predicting human papillomavirus vaccine uptake in young adult women: Comparing the health belief model and theory of planned behavior. *Annals of Behavioral Medicine*, *44*, 171–180.
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*, 562–571.
- Goldstein, N. J., Griskevicius, V., & Cialdini, R. B. (2011). Reciprocity by proxy: A novel influence strategy for stimulating cooperation. *Administrative Science Quarterly*, *56*, 441–473.
- Gollwitzer, P., Sheeran, P., Michalski, V., & Seifert, A. E. (2009). When intentions go public: Does social reality widen the intention-behavior gap? *Psychological Science*, *20*, 612–618.
- Hüffmeier, J., Mazeri, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81–92 (in this issue).
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e24.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.
- Klein, W. M. P., Shepperd, J. A., Suls, J., Rothman, J. A., & Croyle, R. T. (2015). Realizing the promise of social psychology in improving public health. *Personality and Social Psychology Review*, *19*, 77–92.
- Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., ... the Diabetes Prevention Program Research Group (2002n). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine*, *346*, 383–403.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology*, *70*, 556–564.
- Maner, J. K. (2014). Let's put our money where our mouth is: If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives on Psychological Science*, *9*, 343–351.
- Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415–10420.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379–387.
- Mortensen, C. R., & Cialdini, R. B. (2009). Full-cycle social psychology for theory and application. *Social and Personality Psychology Compass*, *4*, 53–63.
- Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*, 574–587.
- Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *3*, 180–183.
- Rand, D. G., Pepsakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014e). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Rinehart & Winston.
- Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science*, *9*, 293–304.
- Sakaluk, J., Williams, A., & Biernat, M. (2014). Analytic review as a solution to the misreporting of statistical results in psychological science. *Perspectives on Psychological Science*, *9*, 652–660.
- Schaller, M. (2016). The empirical benefits of conceptual rigor: Systematic articulation of conceptual hypotheses can reduce the risk of non-replicable results (and facilitate novel discoveries too). *Journal of Experimental Social Psychology*, *66*, 107–115 (in this issue).
- Schaller, M., Park, J. H., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin*, *29*, 637–649.
- Sheeran, P., Abraham, C., & Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: A meta-analysis. *Psychological Bulletin*, *125*, 90–132. <http://dx.doi.org/10.1037/0033-2909.125.1.90>
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *The Robbers Cave experiment: Intergroup conflict and cooperation*. Middletown, CT: Wesleyan University Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.
- Stephens, N. M., Hamadi, M. G., & Destin, M. (2014). Closing the social-class achievement gap: a difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science*, *25*, 943–953.
- Stroebe, W. (2016). Are most published social psychological findings false? *Journal of Experimental Social Psychology*, *66*, 134–144 (in this issue).
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*, 59–71.
- Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science*, *10*, 390–399.