



OPEN ACCESS

EDITED BY

Gideon Nave,
University of Pennsylvania, United States

REVIEWED BY

Benjamin Giguère,
University of Guelph, Canada
Dmitry Grigoryev,
National Research University Higher School of
Economics, Russia

*CORRESPONDENCE

Markus Brauer
✉ markus.brauer@wisc.edu

RECEIVED 24 March 2023

ACCEPTED 19 June 2023

PUBLISHED 23 August 2023

CITATION

Brauer M and Kennedy KR (2023) On effects that do occur versus effects that can be made to occur. *Front. Soc. Psychol.* 1:1193349. doi: 10.3389/frsps.2023.1193349

COPYRIGHT

© 2023 Brauer and Kennedy. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

On effects that do occur versus effects that can be made to occur

Markus Brauer* and Kevin R. Kennedy

Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States

Following Douglas Mook's lead we distinguish between research on "effects that can be made to occur" and research on "effects that do occur" and argue that both can contribute to the advancement of knowledge. We further suggest that current social psychological research focuses too much on the former type of effects. Given the discipline's emphasis on innovation, many published effects are shown to exist under very specific circumstances, i.e., when numerous moderator variables are set at a particular level. One often does not know, however, how frequently these circumstances exist for people in the real world. Studies on effects that can be made to occur are thus an incomplete test of most theories about human cognition and behavior. Using concrete examples, this article discusses the shortcomings of a field that limits itself to identifying effects that might—or might not—be relevant. We argue that it is just as much a scientific contribution to show that a given effect actually does occur as it is to provide initial evidence for a new effect that could turn out to be important. The article ends with a series of suggestions for researchers who want to increase the theoretical and practical relevance of their research.

KEYWORDS

context-specific effects, moderation, theory testing, generalizability, external validity

Introduction

In 1983, Douglas Mook wrote a passionate rebuttal to the then common criticism of the increasingly artificial nature of psychological experiments. Even a finding that does not generalize to the real world can advance knowledge, he argued (Mook, 1983). "One might use the lab not to explore a known phenomenon, but to determine whether such and such a phenomenon exists or *can be made to occur*" [emphasis added] (p. 385). In some studies the researchers' intention is to generalize and to apply their findings to real-life settings. But in other studies, the researchers' goal is to provide an initial test for a mechanism or to "demonstrate the power of a phenomenon by showing that it happens even under unnatural conditions that ought to preclude it" (p. 382). Mook, who talked about all sub-disciplines of psychology, argued that both types of studies contribute to scientific progress.

In the present paper, we build upon Mook's influential article and distinguish between "effects that can be made to occur" and "effects that do occur." We propose that social psychology has turned into a discipline that focuses almost entirely on effects that can be made to occur, which has negatively affected theory development and limited the impact of social psychological research. We begin by defining the difference between effects that can be made to occur and effects that do occur. Then, using concrete examples, we explain how an emphasis on the former has limited the development of social psychological theory. Finally, we outline several measures that researchers can take if their goal is to claim that a certain effect "does occur" and thus has theoretical and practical relevance.

Two different types of effects

Research conducted by social psychologists varies along a continuum (see [Figure 1](#)). On the one end of the continuum are studies that are designed to examine whether certain effects “can be made to occur.” The goal of these studies is to provide initial support for a new idea or to examine a mechanism. They frequently have low external validity, i.e., the findings do not generalize across samples, materials, outcomes, and settings. The methods have been chosen to maximize the likelihood of the effect occurring, which often involves running the study in the laboratory. These studies examine whether something can happen or whether a particular phenomenon is so robust that it can be observed even under circumstances that decrease its likelihood of occurrence. They allow researchers to identify factors that might be relevant for a given phenomenon and thus help them generate predictions for later studies in naturalistic settings.

On the other side are studies that test whether certain effects “do occur.” Here the goal is to determine if a hypothesized effect plays a role in human cognition and behavior in real life-settings.¹ These studies often have high external validity, and the methods are designed to create psychological states that resemble those in real-life settings. The researchers’ intent is to generalize and to make statements about how people, or subgroups of people, think, feel, and behave. These studies examine whether something does happen, whether a particular factor is relevant in the real world, or whether a particular mechanism can be leveraged to change cognitions or behaviors to address social issues.

Note that the distinction between effects that can be made to occur and effects that do occur only marginally overlaps with other dichotomies such as laboratory vs. field research, experimental vs. correlational studies, and basic vs. applied research. Likewise, the distinction is related to but different from the discussions on the lack of representative samples ([Carlson, 1984](#); [Sears, 1986](#); [Henrich et al., 2010](#); [Roberts et al., 2020](#)), the reproducibility crisis ([Pashler and Wagenmakers, 2012](#)), and the utility of social psychological research to help solve societal problems (the “relevance crisis,” [Giner-Sorolla, 2019](#); the “practicality crisis,” [Berkman and Wilson, 2021](#)). As we describe below, it is possible to conduct a reproducible field experiment with a representative sample that examines an effect that can be made to occur. It is also possible for a research program to demonstrate that an effect does occur without contributing much to the field’s understanding of human cognition and behavior in real life-settings.

Effects that can be made to occur

Much of social psychology consists of testing a given theory under specific circumstances. In the typical introduction of a scientific article, the authors (us included) propose theory-derived predictions such as “If theory A is correct, then X should influence Y.” They then report one or more empirical studies with experimental material and procedures that were chosen to maximize the likelihood of finding empirical support for a causal

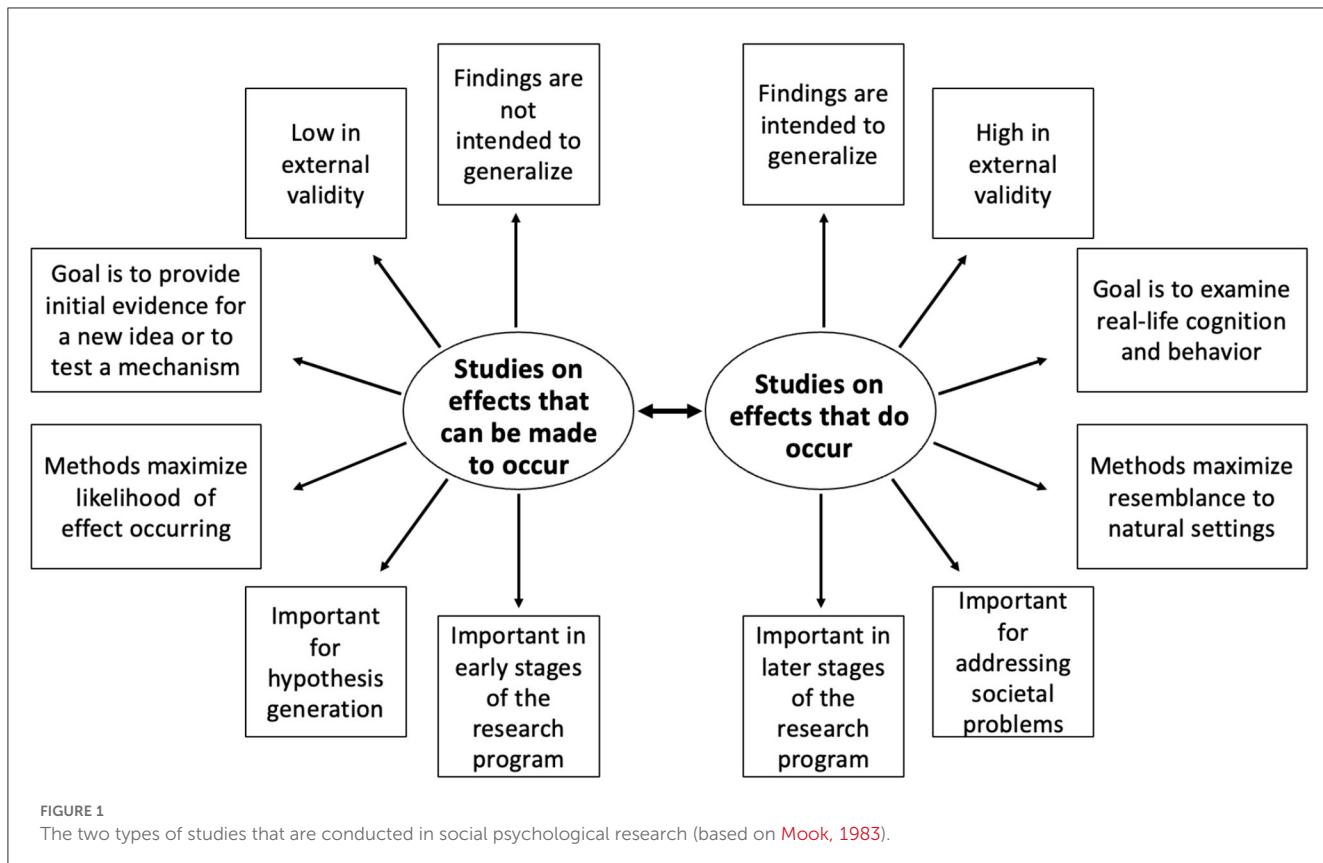
effect of X on Y. Finally, the authors conclude in the General Discussion that they have provided evidence for the idea that X causes Y or, if they are cautious, that X may cause Y. They usually fail to mention that X is likely to cause Y only under the very specific (and maybe even highly unusual) circumstances that they created for their studies. Many published articles show that a given effect can be made to occur, but they do not provide any insight on whether the effect actually does occur with reasonable frequency in people’s daily lives. In other words, these studies are consistent with the proposition that a given effect might play a role in human cognition and behavior, but the reported research does not provide any insight in whether the effect actually does play a role (for a similar claim see [Cialdini, 1980](#); [Baumeister et al., 2007](#); [Mortensen and Cialdini, 2010](#)).

Without a doubt, research on effects that can be made to occur has provided extremely valuable insights and contributed to theory development. Good examples are [Milgram’s \(1974\)](#) landmark studies on obedience and [Asch’s \(1951\)](#) research on conformity. In their daily lives, individuals are rarely asked to shock strangers or judge the length of a line in a group setting. The goal was not to recreate the real world, but instead to examine whether relatively minor forms of social influence or situational pressures can produce consequential behavioral responses. And the finding was “Yes, they can!” More so than anyone would have predicted. In a similar vein, valuable contributions were made by research on priming, salience effects, heuristics, biases, framing effects, the exposure to subliminally presented stimuli, and the influence of emotion states on judgments and decision. It has been shown that subtle changes in wording or question order can affect how people respond on opinion surveys ([Schwarz, 1996](#)) and that unconscious biases can affect our reactions to individuals belonging to certain social groups ([Devine, 1989](#)). As [Prentice and Miller \(1992\)](#) have argued convincingly, it can be important to show that a very minimal manipulation of an independent variable can produce an effect, or that some experimental manipulation can have an impact on a difficult-to-influence dependent variable. It may sound circular, but effects that can be made to occur are interesting when the theoretical question is whether something can happen, rather than whether something does happen ([Mook, 1983](#)). In sum, important scientific progress has been made with studies that examine whether a certain effect can be made to occur, and these studies should continue to have their rightful place in social psychological research.

We suggest, as others have before ([Snyder and Ickes, 1985](#); [Cialdini, 2009](#)), that a discipline can advance knowledge only if it examines both types of effects, those that can be made to occur and those that do occur. For most social psychological theories, it is necessary to examine at some point whether an effect, which has been observed with very specific methods, is found in circumstances that occur with reasonable frequency in people’s daily lives or generalizes across different samples, materials, outcomes, and settings. Without this step, one will never know if the observed effect actually does play a role in human cognition and behavior, which is the primary purpose of most social psychological theories.

The shortcomings of a near-exclusive focus on effects that can be made to occur may not be immediately apparent. We will thus use a concrete example to illustrate it. The article discussed below

¹ We include emotion and motivation in the generic term “human cognition and behavior.”



was chosen because (a) it is highly representative of the kind of studies that we social psychologists conduct, (b) was co-authored by two researchers who have made important contributions to the field, have stellar reputations, and for whom we have the utmost respect, and (c) describes a phenomenon that has been examined in great depth by at least one later publication.

A concrete example

[Song and Schwarz \(2009\)](#) examined if “fluency” can have a causal effect on risk perception. Based on a solid theory that had received ample empirical support in previous years, they predicted that stimuli that can be processed with greater ease will be perceived as entailing fewer risks. They generated 16 ostensible food additives and, based on a pilot study, selected five names that were relatively easy to pronounce and five names that were relatively difficult to pronounce. Participants in the main experiment rated the hazard posed by each of the 10 ostensible food additives. On average, the five difficult-to-pronounce food additives were seen as significantly more harmful than the five easy-to-pronounce ones. The authors replicated this effect in a second study with the same stimuli and using the same within-subjects design. They also conducted a third study in which participants evaluated ostensible amusement park rides and judged the three rides with difficult-to-pronounce names as more adventurous than the three rides with easy-to-pronounce names, regardless of whether a ride being adventurous was presented as a positive (“exciting”) or negative (“risky”) characteristic.

[Song and Schwarz’s \(2009\)](#) research is representative of a common approach to social psychological research: The authors propose a theory-derived prediction and then examine whether the effect can be made to occur. Except for minor methodological shortcomings, the reported studies satisfy the criteria of rigorous science. Their research is of interest for the purpose of the present article because [Bahník and Vranka \(2017\)](#) conducted extensive follow-up research on the same psychological phenomenon and found that the effect depended on particular details of the method.

[Bahník and Vranka \(2017\)](#) replicated the original results when they used exactly the same 10 stimuli as [Song and Schwarz \(2009\)](#) in Study 1. However, when they used a standardized procedure to generate 50 new names for ostensible food additives and used these names as stimuli, there was no relationship between the ease with which the food additives could be pronounced and their perceived harmfulness. Bahník and Vranka found similar null effects when they asked participants to rate the perceived risk of easy- and difficult-to-pronounce criminals, cities in war-stricken Syria, and beach resorts at the Turkish Riviera. It appears that [Song and Schwarz](#) studied an effect that can be made to occur under very specific circumstances, but that goes away when minor, seemingly unimportant elements of the experiment are changed.

When the stars are aligned

Examining the evidence [Song and Schwarz \(2009\)](#) generated in favor of their prediction leads to several important considerations. Like most of us, they used a very specific method (participants,

stimulus material, and procedure). However, from simply reading the article, one does not know which elements of the method are essential for the effect to occur. Song and Schwarz's results can be described as follows (see [Figure 2](#)): Fluency has a causal effect on risk perception if one (a) asks participants to rate multiple food additives, (b) uses the authors' original stimuli (but not other stimuli that were generated with a standardized procedure), (c) gives participants no information except the names of the food additives, (d) uses names that the participants have never heard before, (e) operationalizes fluency through pronounceability, (f) uses a within-subjects design (i.e., exposes participants to both easy-to-pronounce and hard-to-pronounce food additives), (g) uses harmfulness ratings as the indicator for perceived risk, (h) uses college students at a large Midwestern university as participants, and (i) tests participants in a context-less situation in which their responses bear no consequences.

The variables listed in the previous paragraph are typically known as moderators. [Figure 2](#) illustrates that Song and Schwarz set a large number of moderator variables at specific levels, and neither the authors nor the readers know whether the effect persists if one or more of these moderators are set at a different level. [Bahník and Vranka \(2017\)](#) showed that the effect disappears if either (a) or (b) are changed. It may be equally necessary for other moderators to be set at specific levels. For example, the effect may be limited to Western, educated, industrialized, rich and democratic (WEIRD) college students who make judgments about stimuli that they do not care about ([Henrich et al., 2010](#)).

When does fluency affect risk perception? Based on the two articles discussed in the previous paragraphs, a reasonable answer to this question is "When all the stars are aligned" or, said differently, when all moderator variables are set at the "right" levels. We do not know, however, how frequently the stars are aligned for people on this planet during their lifetime. These circumstances could be quite frequent, or they could never occur. Most people might go through life and never be in a situation in which they judge the harmfulness of food additives without any information about the product that contains these additives and without a relevant goal in mind (e.g., protect a family member who has allergies). Thus, although Song and Schwarz have shown that the effect of fluency on risk perception might play a role in risk-related thoughts, judgments, and actions, we do not know if, and when, fluency actually does play a role. Consequentially, the demonstration of the effect is useful in the sense that it provides initial support for a new theoretical proposition but is of limited utility when the goal is to develop comprehensive theories for human cognition and behavior.

The issue is not unethical research practices

Some readers may think that we are suggesting that Song and Schwarz engaged in unethical research practices, which is not at all our proposal. There is nothing wrong with the way Song and Schwarz proceeded in their scientific exploration. They simply did what many of us have done in numerous studies: They generated a theoretically-derived prediction and then developed

an experimental procedure that maximized the chances for the hypothesized effect to occur. This is what we do as scientists.

To illustrate this point, we encourage readers to imagine that their goal is to find out whether it is possible to use a hand saw to cut down a tree with a one-foot diameter trunk. Would one pick an average person and give them the first saw one can find? No, one would probably find a strong logger with extensive experience with hand saws and would give them the biggest and sharpest hand saw that is available. As scientists, we proceed in the same way. If our goal is to examine if a certain effect can be made to occur, we choose a method that will maximize the likelihood that the effect will actually emerge.

[Fiedler \(2011\)](#) described the situation quite accurately: "Researchers [...] select stimuli, task settings, favorable boundary conditions, dependent variables and independent variables, treatment levels, moderators, mediators, and multiple parameter settings in such a way that empirical phenomena become maximally visible. In general, paradigms can be understood as conventional setups for producing idealized, inflated effects" (i.e., effects that can be made to occur, p. 163). He continues: "The selection of stimuli is commonly considered a matter of the researcher's intuition. A 'good experimenter' is one who has a good feeling for stimuli that will optimally demonstrate a hypothetical effect." (p. 164).

A similar point was made by Cialdini nearly 40 years ago when he argued that the work of experimentalists is to "build a sensitive and selective mechanism for snaring the predicted effect" ([Cialdini, 1980](#), p. 23).

[Song and Schwarz \(2009\)](#) would have engaged in unethical research practices if they had conducted multiple pilot studies to identify names of food additives that are either both easy to pronounce and appear to be safe, or both hard to pronounce and appear to be harmful, and if they had then omitted to report these pilot studies in the published article ([Simmons et al., 2011](#)). One can safely assume that this is not how the authors proceeded.

Other ways of doing research are not exactly unethical but are nevertheless problematic. One frequent situation is the one in which the lead researcher and their collaborator(s) generate the stimuli after having formulated the hypothesis. Without being aware and despite their effort to avoid any bias, they may generate a subset of stimuli for which the prediction holds. When Song and Schwarz generated their stimuli, did they unconsciously come up with names that were not only hard to pronounce but that also sounded dangerous? We will never know, but it is interesting that the observed effect disappeared when [Bahník and Vranka \(2017\)](#) used a more standardized procedure to generate easy- and hard-to-pronounce names.

Another potentially problematic practice occurs when the researchers conduct numerous studies and progressively modify the stimulus material and experimental procedure until they get the effect to "work" and subsequently only publish the studies in which the effect was statistically significant. Most of us have heard conference speakers say sentences like "after several trials we finally got the effect that we wanted." Such effects are either type I errors or are limited to conditions in which a large number of moderator variables need to be set at specific levels for the effect to occur ([Button et al., 2013](#)).

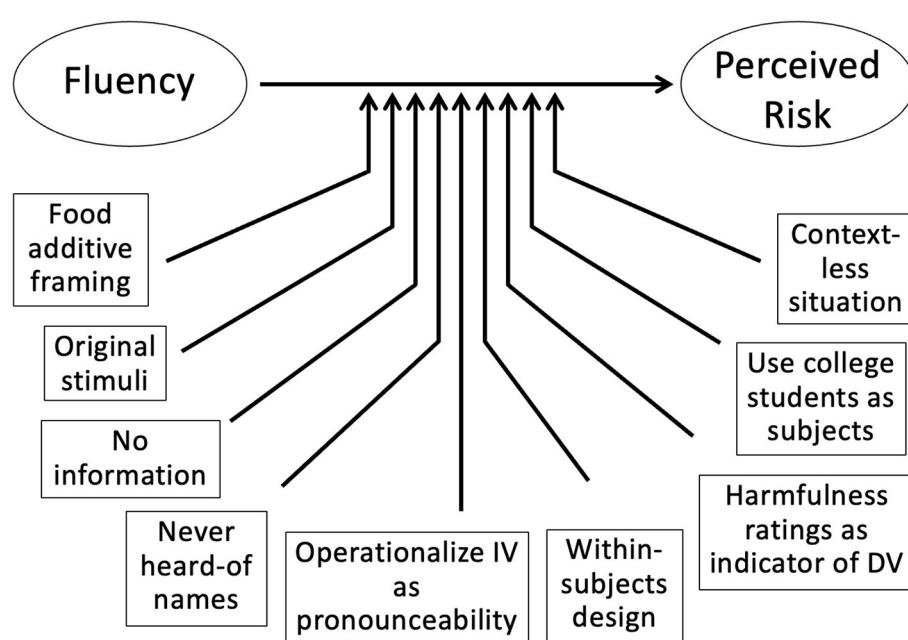


FIGURE 2

Most effects in the literature depend on a large number of moderator variables that are not identified by the authors.

We never make it past the molecule stage

In many scientific fields, researchers start out by studying effects that can be made to occur and then examine whether these effects actually do occur. For example, in pharmacological research, a vast number of potentially interesting effects are initially identified but then only a tiny proportion of them turn out to be relevant for drug development. Every year, researchers create or screen thousands of new molecules with the goal of altering specific biochemical processes in living organisms. Once it has been shown that a molecule can produce a certain effect under highly specific circumstances, it undergoes a series of tests in increasingly complex environments, e.g., lead optimization testing to minimize non-specific interactions, toxicity testing to verify absorption, distribution, metabolism, and excretion, animal testing, low-dose testing on healthy humans, large-scale clinical trials with patients (Hughes et al., 2011). More than 99% of the molecules turn out to be useless: Once they are tested in complex environments, their effects don't hold up for many possible reasons. The effects turn out to be dependent on moderator variables being set at particular levels that do not exist in the real world in that combination, e.g., the effects are eliminated by other substances or the molecules have undesirable side effects (Sun et al., 2022). One way, then, to describe pharmacological research is as follows: Initially, researchers demonstrate a large number of effects that can be made to occur, but subsequent tests reveal that most of them actually do not occur in living organisms (or cannot be leveraged for the treatment of diseases).

Our proposition here is that psychological research frequently does not make it past the molecule stage: We study psychological effects that can be made to occur under specific circumstances,

but we rarely examine if they do occur with reasonable frequency in everyday life or if they can be leveraged to address a societal problem.

Consider a second example from our own area of research on prejudice reduction and intergroup relations. Social psychologists have studied this topic for over 70 years (Dovidio et al., 2010). They have identified numerous effects that can be made to occur. And yet when it comes to giving concrete advice on how to reduce prejudice, social psychologists have little to offer (Paluck et al., 2021). Campbell and Brauer (2020) searched the literature with the goal to identify studies on prejudice reduction that satisfy the following three criteria: (1) The study was a randomized (or a cluster-randomized) experiment, (2) The study assessed outcomes with a delay of at least one month, (3) The study contained "consequential" outcome measures (e.g., grades, drop-out rates, disciplinary actions, turn-over, number of sick days, number of women and minorities in leadership positions, mental or physical health, actual behaviors). They barely found a handful of studies, and even for these, the effect usually emerged for only one of the many outcome measures. Most other prejudice reduction methods have either not been rigorously evaluated or have turned out to be ineffective (Paluck and Green, 2009; Dobbin and Kalev, 2016; Forscher et al., 2019).

There are numerous other examples in the literature. Biased judgments and self-fulfilling prophecies are effects that can be made to occur, but subsequent research has shown that humans are characterized by high levels of rationality and accuracy in daily life (Jussim, 2012). Psychologists have identified numerous processes that might be relevant when the goal is to promote behavior change (e.g., reducing one's carbon footprint, eating healthy), but we generally do not know which of these processes actually do play a role or can be leveraged to produce behavior change. Until recently,

research on behavior change has been dominated by our colleagues in behavioral economics, public health, and social marketing (there are notable exceptions of course, e.g., the research discussed by [Walton and Wilson, 2018](#)).

Given social psychologists' expertise it is surprising that so few of them were asked to join the Social and Behavioral Sciences Team created by President Obama, or the UK's Behavioral Insights Team. In contrast, behavioral economists are often over-represented in these committees.² A priori one might expect federal administrations, businesses, organizations, and the media to have numerous social psychologists in their workforce. And yet, anecdotal evidence suggests that this is not the case ([Brauer et al., 2004](#); [Paluck and Cialdini, 2014](#)). One plausible explanation is that social psychologists are too focused on studying effects that might play a role rather than examining whether they do play a role, limiting the perceived ability of social psychological theory to address societal issues.

To summarize, many social psychologists limit themselves to studying effects that can be made to occur, and there seems to be little appreciation for research that examines whether these effects actually do occur. This focus on "novel, potentially interesting effects" not only restricts the influence of social psychological research beyond its disciplinary boundaries, but it also prevents us from conducting adequate tests of our psychological theories. Note that these theories are supposed to provide abstract descriptions and general laws that govern human cognition and behavior, either for all people or for specific social groups. But given that we frequently limit ourselves to studying processes that might play a role rather than processes that do play a role, it is difficult to draw inferences beyond the specific (and sometimes highly unusual) circumstances in which these processes can be made to occur.

The issue is not reproducibility, field studies, or applied research

The general claim of this paper is that social psychologists frequently limit themselves to studying whether effects can be made to occur under very specific circumstances that maximize the likelihood of occurrence of the effect. As alluded to earlier, this claim is only tangentially related to the discussion on reproducibility that has taken place in recent years. Exact replications, in which the researchers aim to directly follow the original procedure, reveal little as to whether the effect does occur with non-negligible frequency. Even if an effect were 100% reproducible with the same methodology (i.e., under the specific circumstances created by the original authors), we would still not know if the effect actually does play a role in human cognition and behavior. Conceptual replications can provide greater insight

² The Office of Evaluation Sciences (OES), the current iteration of the Social and Behavioral Sciences team, currently employs only one staff member (out of 33) with a doctorate in psychology, despite employing many staff members with doctorates in economics and political science ([Office of Evaluation Sciences, 2023](#)). The lack of social psychologists on such committees is even more surprising given recent criticisms of behavioral economics (see [Maier et al., 2022](#)).

into whether an effect does occur, provided that when designing the replication study the researchers consider the frequency of the "new" circumstances in everyday life (see [Crandall and Sherman, 2016](#) for a further discussion of exact vs. conceptual replications).

There is however one point of contact between our claim and the low reproducibility of psychological findings. When the [Open Science Collaboration \(2015\)](#) published their large-scale replication project and found that only 36% of studies reproduced the original results, some of the authors whose work failed to replicate pointed out that their original procedure had not been used in exactly the same way in the replication study. Often, however, the differences were extremely minor ([Anderson et al., 2016](#)). One may wonder about the utility of scientific findings for theory development when tiny changes in the experimental procedure cause the effects to disappear ([Van Bavel et al., 2016](#)). As [Yarkoni and Westfall \(2017\)](#) put it: "We argue that this is precisely the situation that much of psychology is currently in: elaborate theories seemingly supported by statistical analysis in one dataset routinely fail to generalize to slightly different situations, or even to new samples putatively drawn from the same underlying population" (p. 1107; see also [Yarkoni, 2022](#)).

Although field studies are more likely than laboratory studies to inform us whether a certain effect actually does occur, the proposal of the present paper cannot be reduced to a simple appeal for more field research. Sometimes a field study consists of a demonstration that a certain effect can be made to occur under a limited set of circumstances ([Paluck and Cialdini, 2014](#)). And sometimes a series of laboratory experiments are conducted in a way such that it is obvious that the effect is not conditional on a large number of moderator variables being set at specific levels (see the recommendations outlined below).

In a similar vein, it would be incorrect to say that we are proposing to conduct more applied research. We argue that a near-exclusive focus on effects that can be made to occur poses a threat to fundamental, basic research. Yet, as with the reproducibility issue, our claims are tangentially related to the discussion of the value of basic vs. applied research. Consider the question examined by [Song and Schwarz \(2009\)](#), i.e., whether fluency affects risk perception. Some would argue that it is important to examine people's risk-related decisions in real-life settings (e.g., drunk driving, safe sex, buying insurance, putting guns in lockboxes, evacuating from forest fires) and to examine if people can be nudged into avoiding high-risk behaviors by making the less-risky option easier to process. However, many social psychologists consider these types of applied studies of limited value. It is interesting what Richard Thaler, winner of the Nobel Prize in economics, told conference participants at the 2016 Annual Meeting of the Society for Personality and Social Psychology:

"For some reason [...] applied research has just never gotten the respect in psychology that it has in economics. [...] If you look at all the economists that have won Nobel Prizes and various other awards, many of them are doing what looks like very applied research. [...] Applied research has always been held in high regard in economics and it's up to you to change that if you want your field to be relevant." ([Thaler, 2016](#)).

It is likely that applied research on variables affecting risk perception would not only help people make better risk-related decisions but would also speak to the number of moderator variables that have to be set at the “right” levels before the effect can be made to occur. Similar arguments can be made about recent proposals on the importance of “scaling up” psychological effects (Oyserman et al., 2018) and conducting intervention research (Walton and Crum, 2020). When psychological effects are examined on a large scale in the field, it is virtually impossible to limit the implementation to extremely specific circumstances. As such, intervention research contributes to establishing the generalizability of an effect, i.e., it speaks to whether a given effect does occur.

Recommendations

What kind of evidence is necessary before one can claim that a given effect actually does occur? What can social psychologists do to conduct research that better informs general theories of human cognition and behavior in everyday settings? What changes can we implement as a field to promote research on effects that do occur? We propose some suggestions in the following sections (see Table 1 for a summary). We put them in the order in which they become relevant for researchers as their research program unfolds and we grouped them into three categories: methodological choices, stimulus selection and analysis, and the way we communicate about research. Many suggestions, especially those in the first category, are not new and have been proposed by others.

Methodological choices

Researchers should select a sufficiently large sample of participants from the entire population of individuals to which they want to generalize their effects. Ideally, this selection is random (Brewer and Crano, 2014). If this is not possible, then at least the researchers should show the effect with more than one participant sample. For example, if they use undergraduates in one study, then they should demonstrate that the effect replicates with at least one sample that does not consist of undergraduates.

A similar reasoning applies to stimuli in that it is desirable to provide evidence for the hypothesized effect with more than one stimulus sample. If participants evaluated cars in Study 1, then researchers may ask Study 2 participants to evaluate abstract paintings. If one of the studies was about doctors, then the next study can be about lawyers.

Researchers need to show a given effect multiple times, each time with a different operationalization of the “causal construct” (the independent variable). The variability can be achieved by changing the format or mode of presentation of the stimuli, the way a particular psychological state is induced, the procedure, the setting, or the cover story. The causal construct can sometimes be measured and other times be manipulated (Wilson et al., 2010). If the independent variable varied within-subjects, then it could be helpful to show that the effect also emerges in a between-subjects design, and vice-versa (Smith, 2014). If different operationalizations of the causal construct were used across studies,

TABLE 1 A list of research practices that researchers can adopt when their goal is to show that a certain effect actually does occur and does play a role in everyday human cognition and behavior.

1. Use large samples that are representative of the population of individuals to which you want to generalize. Vary participant samples between studies.
2. Use numerous stimuli that are representative of the population of stimuli to which you want to generalize. Vary stimulus samples between studies. Use a standardized procedure to generate new stimuli.
3. Across studies use different operationalizations of the “causal construct” (the independent variable). Vary whether this construct is measured or manipulated and whether it varies within or between participants.
4. Across studies use different operationalizations of the “effect construct” (the dependent variable). If relevant, assess outcomes with a delay and use behavioral indicators of these outcomes.
5. Consider including a study in your research program where participants are not aware that they are part of a research study (when feasible).
6. Avoid superficial processing of the experimental material by making sure that participants’ answers and reactions matter for them or for others.
7. Show your effect at least once in a field context. Change settings between studies to make sure the effect is not limited to one context with very specific characteristics.
8. In the statistical analyses, treat stimuli as a random variable. If you decide to treat stimuli as a fixed variable, state specifically in the General Discussion that your results may not generalize beyond the stimuli used in your study.
9. State specifically in your manuscript whether your study was designed with the goal to generalize the findings beyond the sample, materials, and setting used in your study.
10. In the General Discussion reflect on the likelihood that individuals in the real world will experience similar psychological states and constraints as in your studies.
11. Add a section on “constraints on generality” in your General Discussion.
12. Provide arguments or data supporting the idea that the particular levels of the moderator variables you chose for your research occur with reasonable frequency in the real world.

it may be useful to demonstrate that the same psychological construct was measured/manipulated in all the studies.

In the same vein, a finding is more likely to be an effect that does occur if it has been shown to exist with multiple indicators of the “effect construct” (the dependent variable). If the authors used self-reports in one study, they may consider using reaction times or peer ratings in the next study. If the independent variable and the dependent variable were both measured then they may be correlated only because they were assessed with the same method (e.g., self-reports; Podsakoff et al., 2003). It is thus important to show, in a second study, that the correlation persists if the two variables are assessed with different methods.

Depending on the nature of the scientific question under investigation, one of the indicators should be a behavioral outcome measure and one of the studies should assess outcomes with a delay (assuming that the theory predicts a certain longevity of the hypothesized effect). The gold standard for effects that do occur are studies where participants are unaware that they are part of a psychological experiment at the time the outcome variable is assessed, but there are numerous research questions for which such an approach is not feasible for ethical or logistical reasons.

In many psychological experiments, participants complete tasks that they do not care about and where there are no

consequences for “bad” choices or superficial processing of the stimulus material. A certain effect may emerge when undergraduates complete a mock hiring task for course credit, but this same effect may play no role when faculty select new graduate students or when hiring managers select new employees. [Bicchieri \(2017\)](#) suggests incentivizing participants for “good answers” in at least some of the studies. Another way to increase task relevance is to make participants believe that they will work with the person they evaluate most positively or that their answers will influence important decisions that the university leadership will make in the near future.

Researchers need to show awareness of the role of context in theory development. It has been shown that context and the greater environment often are strong moderators of observed effects ([Oishi, 2014](#); [Walton and Yeager, 2020](#); [Berry, 2022](#)). It is therefore desirable for researchers to show the effect at least once in a field context in which participants’ responses have meaning and tangible consequences for them. Note that “field context” is not necessarily equivalent to applied research deprived of theory. Numerous researchers have provided convincing arguments for the usefulness of field research for the development of psychological theories ([Paluck and Cialdini, 2014](#); [Power and Velez, 2022](#)). In a recent review, [Mitchell \(2012\)](#) found that 20% of the social psychological effects shown in a laboratory setting change signs when tested in a field setting. Such findings indicate a greater need to understand whether effects that can be made to occur do, in fact, occur in real-life settings.

Another good practice is to implement conceptual replication studies in which the researchers change the settings of one or more experimental design choices that are highly artificial yet appear necessary for the effect to occur. If an effect is found in a setting that is devoid of contextual information and unrepresentative of any situation people might encounter in the real world, it might be helpful to show that the effect also occurs when participants are provided with richer information. For example, if a critical reviewer remarked that an observed effect was likely due to the fact that participants knew they were being filmed, it is helpful to replicate the effect in a more natural setting in which participants’ behaviors are observed unobtrusively.

It may be virtually impossible to achieve all the desirable characteristics listed above in a single study or even in a single scientific article with multiple studies. However, few of us publish a single article on a topic and then move on to explore an entirely different theoretical question. Most of us conduct research programs that span over several years and sometimes even decades. In the course of such a research program, it is quite possible to design and implement studies that possess the above-mentioned characteristics.

Numerous authors have proposed that different types of validity vary in importance in different stages of a research program ([Hoyle et al., 2001](#)). Initially, it may be important to show that a certain effect can be made to occur. In later stages, we propose, it is necessary to show that the effect actually does occur with reasonable frequency in real-life settings. In other words, researchers need to show that the effect generalizes across different participant samples, different stimulus samples, different

operationalizations of the independent variable(s), different outcome measures, different contexts, and that it occurs in settings that people might encounter with non-negligible frequency in the real world.

Selecting stimuli and treating stimuli as a random factor

Most research studies require participants to provide a behavioral response to one or more stimuli. The behavioral response can be a click on a multi-point Likert scale in a Qualtrics survey, an eye-movement, a hiring decision, or any other form of verbal or non-verbal behavior. The stimuli can be words, pictures of individuals, vignettes, images, videos, confederates, target groups (e.g., thermometer ratings), job applications, and the like. Here, we will discuss one aspect of methodological decisions that warrants its own section: the impact of different ways of selecting and analyzing stimuli on one’s ability to conclude that a certain effect does occur.

Stimulus selection is crucial. If researchers want to claim that their effects generalize beyond the specific stimuli used in their study, it is necessary to randomly select the stimuli from the pool of available stimuli. Several authors have insisted that we should be thinking about stimuli the same way we have always thought about participants: We randomly select them from the population to which we want to generalize and the sample of stimuli should be sufficiently large ([Clark, 1973](#); [Westfall et al., 2014](#)). For example, just as we would never consider recruiting only four participants, we should never run a study with only four stimuli.

If researchers use the headshots of two Black and two White individuals as stimuli in a study they know that the observed effect can be made to occur with these four target individuals. If, however, they (randomly) select headshots among a large pool of Black and White individuals, it is considerably more likely that the effect actually does occur, especially if the sample of headshots is large. The research on the “risky shift phenomenon” is an example where numerous authors drew incorrect conclusions simply because they all used the same stimulus set with unusual characteristics ([Westfall et al., 2015](#)).

If it is impossible to have participants react to a large number of stimuli, it may be possible to expose each participant to a small number of randomly selected stimuli from the entire pool of stimuli (i.e., each participant sees their own set of randomly selected stimuli).

If researchers want to use artificial or new stimuli, there should be a (standardized) procedure for generating them. At the minimum, the stimuli should be selected by individuals who are unaware of the hypothesis. Stimuli should also be pilot-tested to make sure that they vary only on the to-be-manipulated dimension and not on other subjective dimensions (e.g., likeability, ease of processing). Finally, researchers should examine whether stimuli differ on irrelevant dimensions that can be measured objectively (e.g., number of letters in target words, number of words in vignettes, lighting of headshots).

When the same group of participants judge the same set of stimuli, researchers have the choice between treating stimuli as a

fixed variable or as a random variable.³ Treating stimuli as a fixed factor usually leads to increased statistical power (smaller standard errors), but researchers should be aware that by doing so their results cannot be generalized beyond the stimuli used in the study. [Song and Schwarz \(2009\)](#) treated the 10 food additives they used in their study as a fixed factor, but failed to state explicitly that this data-analytic choice prevented them from drawing conclusions about easy- and hard-to-pronounce food additives in general.⁴ To the contrary, the authors seem to imply, in the General Discussion, that their findings have implications for the management of perceived risk in applied settings, a conclusion unwarranted by the data analytic strategy.

[Judd et al. \(2012\)](#) showed that treating stimuli as a fixed factor leads to increased type I error rates when the researchers' goal is to generalize their findings to the entire pool of stimuli from which the experimental stimuli were drawn. Given that [Song and Schwarz](#) used 10 stimuli and 20 participants, their type I error rate was approximately 30% (assuming that their goal was to be able to generalize the observed effects beyond the 10 names used in the study). It is not surprising, then, that the effects did not hold up when [Bahník and Vranka \(2017\)](#) used comparable but different stimuli.

Researchers often want to generalize their findings beyond the specific stimuli used in the study. In that case, they have no choice but to treat both participants and stimuli as random factors. Several articles have been written on this topic, and most of them provide researchers with the syntax to conduct the appropriate linear mixed-effects models ([Judd et al., 2017](#); see [Brauer and Curtin, 2018](#)).

According to [Yarkoni \(2022\)](#), most psychological research projects are intended to generalize across design elements other than just participants and stimuli. Usually our goal is to generalize across tasks, instructions, experimenters, and research sites. If such generalization is crucial to the theory then these other design elements should also be treated as random factors. One possibility is to focus on a subset of design elements in each study. For example the first study could focus on participants and stimuli. Both participants and stimuli are randomly selected from the pool of individuals and the pool of stimuli to which the researcher wishes to generalize, and both participants and stimuli are treated as random factors. The second study could then focus on participants and tasks, i.e., both participants and tasks are randomly selected from the pool of individuals and the pool of tasks to which the researcher wishes to generalize, and both participants and tasks are treated as random factors (see [Yarkoni, 2022](#), for other suggestions).

Communicating research results

We suggest writing and talking about research in a way that clearly distinguishes between effects that can be made to occur and effects that do occur. Researchers should clearly articulate to

³ Note that participants are always treated as a random variable.

⁴ Admittedly [Song and Schwarz's](#) paper was published three years before [Judd et al.'s \(2012\)](#) influential article that raised awareness about the consequences of treating stimuli as fixed vs. random factors.

the reader whether the goal of the research project was to provide initial evidence for a hypothesized effect in circumstances that maximized the likelihood of its occurrence, or whether the goal was to demonstrate the existence of an effect in circumstances that people encounter with appreciable frequency in the real world.

One solution is to carefully choose the formulation used in the General Discussion of one's article. When describing studies that were designed to show that a certain effect can be made to occur, authors should avoid sentences such as "we have shown that X causes Y." More appropriate ways to describe such scientific contributions are "X may influence Y" and "we have shown that under the certain circumstances described here X can influence Y" (see [Yarkoni, 2022](#), for a similar claim).

We also encourage researchers to reflect on the likelihood that the chosen combination of moderator variables is one that occurs at least occasionally in real-life settings. [Aronson and Carlsmith \(1968\)](#) argued that an experimental procedure can be artificial but what is key is whether it induces the psychological state that people might experience in real life situations. Authors may want to comment on the extent to which participants were induced to experience psychological states that are representative of real-world settings. Further, authors should consider how the reader will interpret the theoretical findings of the experiment. Even if the authors do not intend for their findings to generalize beyond the confines of their study, the reader may be tempted to draw unwarranted conclusions. A more explicit acknowledgment of effects that can occur vs. do occur will provide greater context for the reader's understanding of the results and downstream implications.

We also concur with [Simons et al. \(2017\)](#) who suggested including a statement on "Constraints on Generality" in the General Discussion of scientific articles. Such a statement "explicitly identifies and justifies the target populations for the reported findings" (p. 1124). As psychologists, we use samples to draw inferences about populations of people, situations, and stimuli. Simons and colleagues suggest that article authors name the target populations to which they would like to generalize their findings and discuss the aspects of the studies that limit their capacity to do so. Further, [Roberts et al. \(2020\)](#) recommend specifically detailing the demographics of samples and [Berry \(2022\)](#) has proposed that authors write a more explicit acknowledgment and description of the cultural and ecological influences that affect participant's behavior within a study.

Researchers can go further and specifically provide arguments or (even better) data explaining that the particular levels of the moderator variables they chose for their research occur with reasonable frequency in the real world. For example, consider a study that shows that making a particular idea highly salient causes participants to interpret the ambiguous behavior of a target person in a manner consistent with the idea. Researchers could then discuss whether such extreme levels of salience can naturally occur, and how frequently individuals in the real world find themselves in a situation where they try to interpret ambiguous behavioral information without context and without any consequences for incorrect interpretations. They could also conduct a follow-up study in which they ask individuals to report how frequently in the last 5 years they found themselves in situations that are psychologically comparable to the one in the experiment.

Conclusion

The emphasis on effects that can be made to occur vs. effects that do occur reflects what is valued by the field. Studies demonstrating that an effect can be made to occur play a key role for theoretical innovation and can provide valuable insights. As mentioned above, they can identify an important mechanism, speak to the robustness of an effect, or demonstrate that people's scores on a difficult-to-change outcome variable can be modified by an experimental manipulation. But if researchers want to make general claims about human cognition and behavior or even address societal problems, it will be necessary that they show at some point in their research program that the effect actually does occur. Doing that requires more than showing that a manipulated X influences Y under specific circumstances that have been created to maximize the likelihood of occurrence of the effect.

Researchers adopting different research practices, even if well intended, is likely to be insufficient. Changes in research and publication practices occur only if the incentive structure changes. For example, several years ago, the editorial team of the *Personality and Social Psychology Bulletin* encouraged authors to include in the General Discussion a section on "Constraints on Generality" modeled after the proposal by Simons et al. (2017) mentioned above. It also encouraged conceptual replications that "extend operationalizations and test theories in new ways" (Crandall et al., 2018, p. 288). These were excellent steps in the right direction.

Still, further changes are warranted. Editors and reviewers should agree that a series of studies showing that a previously observed effect does occur and plays a meaningful role for humans in real life is as much a scientific contribution as a study that provides initial evidence that a novel effect can be made to occur. To borrow the metaphor of pharmacological research one more time: It is interesting to show that a certain new molecule can establish a chemical reaction in vitro. But it is an equally important scientific contribution to show that the molecule has the desired effect in vivo (and that it can be used in the treatment of diseases). When authors cite evidence for a certain effect, they may decide to cite the article that first showed that the effect actually does occur, rather than the article that demonstrated that the effect can be made to occur. A viable compromise would be to cite both articles.

Variation of experimental methods from one study to the next, even within the same manuscript, should be considered a strength rather than a weakness. The demonstration that a certain effect holds across different levels of important moderator variables—that the effect is not limited to very specific, possibly artificial circumstances—should increase a manuscript's likelihood of acceptance. Likewise, if a team of researchers has published numerous articles showing that a certain effect can be made to occur in circumstances that are unlikely to occur in the real world,

it is acceptable for a reviewer to request evidence that the effect actually does occur before another manuscript on this effect can be accepted for publication.

The key proposition of this article is that we should stop focusing exclusively on novel effects that can be made to occur under very specific circumstances. Although research on these types of effects has its rightful place in our scientific endeavors, it should be complemented by research on effects that do occur. Only by studying effects that generalize across samples, stimuli, outcomes, and settings will we be able to provide scientific evidence for general theories on human cognition and behavior and identify mechanisms that can be leveraged to induce positive societal change. We entirely agree with Mook (1983) when he says that "ultimately, what makes research findings of interest is that they help us understand everyday life" (p. 386). This understanding does not just come from studies on what can happen but also from research on what does happen in the real-world.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MB wrote the first draft of the manuscript. MB and KK substantially revised and edited the manuscript and approved the final manuscript for submission. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderson, C. J., Bahnik, Š., Barnett-Cowan, M., Bosco, F. A., and Chandler, J. (2016). Response to a comment on "Estimating the reproducibility of psychological science". *Science* 351, 1037. doi: 10.1126/science.aad9163

Aronson, E., and Carlsmith, J. M. (1968). "Experimentation in social psychology." In *Handbook of Social Psychology*, eds G. Lindzey and E. Aronson, 2nd Edn (London: Addison-Wesley), 1–79.

- Asch, S. E. (1951). *Effects of Group Pressure Upon the Modification and distortion of Judgments*. New York, NY: Carnegie Press.
- Bahnik, Š., and Vranka, M. A. (2017). If it's difficult to pronounce, it might not be risky: the effect of fluency on judgment of risk does not generalize to new stimuli. *Psychol. Sci.* 28, 427–436. doi: 10.1177/0956797616685770
- Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Persp. Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x
- Berkman, E. T., and Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. *Persp. Psychol. Sci.* 16, 864–874. doi: 10.1177/1745691620969650
- Berry, J. W. (2022). The forgotten field: Contexts for cross-cultural psychology. *J. Cross-Cult. Psychol.* 53, 993–1009. doi: 10.1177/00220221221093810
- Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.
- Brauer, M., and Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: a unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychol. Methods* 23, 389–411. doi: 10.1037/met0000159
- Brauer, M., Martinot, D., and Ginet, M. (2004). Current tendencies and future challenges for social psychologists. *Curr. Psychol. Cognit.* 22, 537–558.
- Brewer, M. B., and Crano, W. D. (2014). "Research design and issues of validity," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis and C. M. Judd, 2nd Edn (Cambridge: Cambridge University Press), 11–26.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Campbell, M. R., and Brauer, M. (2020). Incorporating social-marketing insights into prejudice research: advancing theory and demonstrating real-world applications. *Persp. Psychol. Sci.* 15, 608–629. doi: 10.1177/1745691619896622
- Carlson, R. (1984). What's social about social psychology? Where's the person in personality research? *J. Pers. Soc. Psychol.* 47, 1304–1309. doi: 10.1037/0022-3514.47.6.1304
- Cialdini, R. B. (1980). "Full-cycle social psychology," in *Applied Social Psychology Annual*, eds L. Bickman (London: Sage Publications), 21–47.
- Cialdini, R. B. (2009). We have to break up. *Persp. Psychol. Sci.* 4, 5–6. doi: 10.1111/j.1745-6924.2009.01091.x
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verb. Learning Verb. Behav.* 12, 335–359. doi: 10.1016/S0022-5371(73)80014-3
- Crandall, C. S., Leach, C. W., Robinson, M., and West, T. (2018). PSPB editorial philosophy. *Pers. Soc. Psychol. Bullet.* 44, 287–289. doi: 10.1177/0146167217752103
- Crandall, C. S., and Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *J. Exp. Soc. Psychol.* 66, 93–99. doi: 10.1016/j.jesp.2015.10.002
- Devine, P. G. (1989). Stereotypes and prejudice: THEIR automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5
- Dobbin, F., and Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review* 94, 52–60.
- Dovidio, J. F., Hewstone, M., Glick, P., and Esses, V. M. (2010). *The SAGE Handbook of Prejudice, Stereotyping and Discrimination*. London: Sage Publications Inc.
- Fiedler, K. (2011). Voodoo correlations are everywhere—Not only in neuroscience. *Persp. Psychol. Sci.* 6, 163–171. doi: 10.1177/1745691611400237
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., et al. (2019). A meta-analysis of procedures to change implicit measures. *J. Pers. Soc. Psychol.* 117, 522–559. doi: 10.1037/pspa0000160
- Giner-Sorolla, R. (2019). From crisis of evidence to a "crisis" of relevance? Incentive-based answers for social psychology's perennial relevance worries. *Eur. Rev. Soc. Psychol.* 30, 1–38. doi: 10.1080/10463283.2018.1542902
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X
- Hoyle, R. H., Harris, M., and Judd, C. M. (2001). *Research Methods in Social Relations*. New York, NY: Thomson Learning.
- Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011). Principles of early drug discovery. *Br. J. Pharmacol.* 162, 1239–1249. doi: 10.1111/j.1476-5381.2010.01127.x
- Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *J. Pers. Soc. Psychol.* 103, 54–69. doi: 10.1037/a0028347
- Judd, C. M., Westfall, J., and Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Ann. Rev. Psychol.* 68, 601–625. doi: 10.1146/annurev-psych-122414-033702
- Jussim, L. (2012). *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-Fulfilling Prophecy*. Oxford: Oxford University Press.
- Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., and Wagenmakers, E. J., et al. (2022). No evidence for nudging after adjusting for publication bias. *Proc. Nat. Acad. Sci.* 119, e2200300119. doi: 10.1073/pnas.2200300119
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Harper and Row.
- Mitchell, G. (2012). Revisiting truth or triviality: the external validity of research in the psychological library. *Persp. Psychol. Sci.* 7, 109–117. doi: 10.1177/1745691611432343
- Mook, D. G. (1983). In defense of external invalidity. *Am. Psychol.* 38, 379–387. doi: 10.1037/0003-066X.38.4.379
- Mortensen, C. R., and Cialdini, R. B. (2010). Full-cycle social psychology for theory and application. *Soc. Pers. Psychol. Compass* 4, 53–63. doi: 10.1111/j.1751-9004.2009.00239.x
- Office of Evaluation Sciences (2023). *Team*. Available online at: <https://oes.gsa.gov/team/> (accessed June 8, 2023).
- Oishi, S. (2014). Socioecological psychology. *Ann. Rev. Psychol.* 65, 581–609. doi: 10.1146/annurev-psych-030413-152156
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349, aac4716. doi: 10.1126/science.aac4716
- Oyserman, D., Yan, V., and Lewis, N. (2018). *Social and Personality Psychology of Scaling Up: What do we mean by scaling up?* Presentation given at the General Meeting of the Society for Personality and Social Psychology, Atlanta, GA (March 2018).
- Paluck, E. L., and Cialdini, R. B. (2014). "Field research methods," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis and C. M. Judd (Cambridge: Cambridge University Press), 81–97.
- Paluck, E. L., and Green, D. P. (2009). Prejudice reduction: what works? A review and assessment of research and practice. *Ann. Rev. Psychol.* 60, 339–367. doi: 10.1146/annurev-psych.60.110707.163607
- Paluck, E. L., and Porat, R., Clark, C. S., and Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Ann. Rev. Psychol.* 72, 533–560. doi: 10.1146/annurev-psych-071620-030619
- Pashler, H., and Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? [Editorial]. *Persp. Psychol. Sci.* 7, 528–530. doi: 10.1177/1745691612465253
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J. Appl. Psychology* 88, 879–903. doi: 10.1037/0021-9010.88.5.879
- Power, S. A., and Velez, G. (2022). Field social psychology. *Am. Psychol.* 77, 940–952. doi: 10.1037/amp0000931
- Prentice, D. A., and Miller, D. T. (1992). When small effects are impressive. *Psychol. Bulletin* 112, 160–164. doi: 10.1037/0033-295X.112.1.160
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., and Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Persp. Psychol. Sci.* 15, 1295–1309. doi: 10.1177/1745691620927709
- Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. London: Lawrence Erlbaum Associates.
- Sears, D. O. (1986). College sophomores in the laboratory: influences of a narrow data base on social psychology's view of human nature. *J. Pers. Soc. Psychol.* 51, 515–530. doi: 10.1037/0022-3514.51.3.515
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632
- Simons, D. J., Shoda, Y., and Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Persp. Psychol. Sci.* 12, 1123–1128. doi: 10.1177/1745691617708630
- Smith, E. R. (2014). "Research design," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis and C. M. Judd (Cambridge: Cambridge University Press), 27–48.
- Snyder, M., and Ickes, W. (1985). "Personality and social behavior," in *Handbook of Social Psychology*, eds G. Lindzey, and E. Aronson (New York, NY: Random House), 883–948.
- Song, H., and Schwarz, N. (2009). If it's difficult to pronounce, it must be risky: fluency, familiarity, and risk perception. *Psychol. Sci.* 19, 135–138. doi: 10.1111/j.1467-9280.2009.02267.x
- Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sinica B* 12, 3049–3062. doi: 10.1016/j.apsb.2022.02.002

- Thaler, R. (2016). *Improving public policy: How psychologists can help*. Presentation given at the General Meeting of the Society for Personality and Social Psychology, San Diego, CA (January 2016).
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., and Reiner, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proc. Nat. Acad. Sci.* 113, 6454–6459. doi: 10.1073/pnas.1521897113
- Walton, G. M., and Crum, A. J. (2020). *Handbook of Wise Interventions*. New York, NY: Guilford Publications.
- Walton, G. M., and Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review* 125, 617–655. doi: 10.1037/rev0000115
- Walton, G. M., and Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science* 29, 219–226. doi: 10.1177/0963721420904453
- Westfall, J., Judd, C. M., and Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science* 10, 390–399. doi: 10.1177/1745691614564879
- Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General* 143, 2020–2045. doi: 10.1037/xge0000014
- Wilson, T. D., Aronson, E., and Carlsmith, K. (2010). The art of laboratory experimentation. In S.T Fiske, D.T. Gilbert, and G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 51–81). John Wiley and Sons, Inc. doi: 10.1002/9780470561119.socpsy001002
- Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences* 45, 1–78. doi: 10.1017/S0140525X20001685
- Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Persp. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393