Journal of Experimental Social Psychology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Commentary: A big problem requires a foundational change

John F. Dovidio

Yale University, United States

ARTICLE INFO

Available online xxxx

ABSTRACT

This commentary discusses specific insights offered in each article but also attempts to integrate the various contributions to the Special Issue around a tension in the field between increasing certainty of knowledge and the rapid advancement of knowledge that has characterized social psychology historically. While addressing the replicability problem can be done incrementally, fundamental change is needed in the reward structure of the profession, and therefore in the publication system that is the gatekeeper for scholarly success. The commentary considers ways to ensure the stability and robustness of published findings in social psychology while still fostering creative, often counterintuitive, discovery – a hallmark of social psychology – in a scientifically responsible way. Concrete suggestions to encourage a balance in the values of scientific certainty and scientific discovery through the publication system are proposed.

1. Introduction

This Special Issue of the *Journal of Experimental Psychology* is dedicated to presenting practical methodological solutions to the problem of replicability in social psychology. It includes articles in which leading scholars offer a range of perspectives on the "replicability problem" and suggest concrete interventions to address this issue. In this commentary, I discuss, selectively, specific insights offered in each article but also attempt to integrate the various contributions around a fundamental tension between increasing certainty of knowledge and the rapid advancement of knowledge that has characterized social psychology historically. This tension is not unique to social psychology; it likely applies to other subdisciplines of psychology and to other scientific fields. Nevertheless, there are current circumstances, historical influences, and structural factors that make this an acute and timely issue in social psychology.

Crandall and Sherman observe in their article that "science is a social phenomenon." Thus, while this Special Issue includes proposals for specific solutions to the problem of replicability in social psychology, consideration of the replicability problem requires broader discussion of the larger social and professional research context. For a substantial number of individuals in the profession, research is not simply an activity; *publication* of one's research activity is essential for retaining one's job (through tenure decisions), academic advancement, prestige, and public as well as professional recognition. Jussim, Crawford, Anglin, Stevens, and Duarte (in press), in their article of this Special Issue note:

> Most of us are motivated to get the science right, but we are also motivated to get the studies published and our grants funded. We want our colleagues to find our research sufficiently interesting and important to support publishing it, and then to cite it, preferable a lot. We want jobs, promotion, and tenure. We want popular media to publicize our research and to disseminate our findings.

The publication system thus shapes research activity and, ultimately, the scientific values and standards in the field.

While similar publication pressures exist in other fields, there may be distinctive contemporary and historical forces in social psychology. One of these factors is the current competitiveness of social psychology journals. The most respected outlets in social psychology, including the *Journal of Experimental Social Psychology*, have rejection rates between 75% and 90%. This rejection rate far exceeds those in many physical sciences (e.g., in atmospheric sciences, around 35%; Schultz, 2010), and is generally higher than in allied subdisciplines and disciplines.

As a consequence of a publication system that is perceived to separate the "great from the near-great," authors consistently seek new ways to maximize their chances for publication. For example, Reis and Stiller (1992) summarized publication trends in the *Journal of Personality and Social Psychology*, noting that in more recent years articles were more complex and contained more studies. Reis and Stiller also mentioned, as a seemingly minor point, that over time more published articles included colons in their title. Although the causal connection cannot be established methodologically, it is interesting to note that in the two years prior to that article, fewer than half of the articles published in the *Journal of Experimental Social Psychology* had titles with colons (46% in 1990 and 43% in 1991). Within a few years of that publication, 58% of the titles did (60% in 1997 and 56% in 1998). In 2013, 65% of the titles in the journal included colons.

E-mail address: john.dovidio@yale.edu.

http://dx.doi.org/10.1016/j.jesp.2016.01.008
0022-1031/

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

The motivation of authors to gain a publication edge can have more substantive effects on social psychology and the field more generally. As Hales (in press) writes in his article for this Special Issue:

> It has been observed that individual researchers are motivated to seek truth, but are also motivated to amass publications for career advancement. To the extent that the most efficient ways to publish articles do not promote truth (because questionable research practices enhance the chances of publication), individuals are tempted to show reduced concern for the truth of published finding (Nosek & Bar-Anan, 2012; Nosek, Spies, & Motyl, 2012).

Perceived pressure to publish may be a factor in extreme cases, such as data fabrication (Stroebe, Postmes, & Spears, 2012). Such cases are immensely damaging to the profession, but they are only one, presumably restricted, facet of the replicability problem. Another, more likely widespread consequence is p-hacking, in which researchers engage in a number of different "data massaging" strategies to obtain the desired result (Simmons, Nelson, & Simonsohn, 2011). Many of these kinds of practices, such as reporting on a subset of dependent measures, running more participants until a marginally significant effect has attained "$p < .05$," or failing to report nonreplications (see Francis, 2012; Schimmack, 2012), had been relatively common practices in the field in the past, not fully recognized as improper or even as "gray areas." However, as newer articles (including those in the present Special Issue) explain, such practices produce "false positives" and directly contribute to the replicability problem. Head, Holman, Lanfear, Kahn, and Jennions (2015) identified psychology as one of the disciplines in which p-hacking is particularly problematic.

The issue of how common practices in social psychology produce false positives is a critical one today, but it is not a new problem. Because of its roots in strong motivations to publish, it has been longstanding issue. Forty years ago, in a comprehensive article describing "prejudice against the null hypothesis and its consequences," Greenwald (1975) surveyed researchers and reviewers to identify attitudes, orientations, and practices in psychology that have system-wide effects inflating the proportion of published findings representing Type I errors. He then modeled the research and publication process and warned that almost a third (30%) of published articles could represent Type I errors. Greenwald concluded that the attitudes and behavior of researchers play a major role in the problem of the publication of Type I errors: "the problem exists as much or more in the behavior of investigators before collecting and after analyzing their data as in the techniques they use for their analysis" (p. 16). The problem of false positives has thus been an enduring issue in social psychology.

A second social factor that has affected the science of social psychology is its historical academic position. Social psychology has long been a tenuous science. It represents a gateway subdiscipline of psychology, bridging micro-levels of analysis (e.g., brain imaging drawing on techniques from human neuroscience) with macro-levels of analysis (e.g., structural and ideological processes, drawing from work in political science and sociology). This strong and unique emphasis on the influence of social context has raised questions about whether it can identify the kinds of invariant relationships that characterize work in the physical and biological sciences. Indeed, a prominent critique entitled, "Social Psychology as History," Gergen (1973) questioned whether social psychology should even be considered a science.

Perceptions of the tenuous position of social psychology as a science have contributed to a series of "crises" over the years, including questions about social psychology's social relevance (Rodin, 1985) and challenges relating to its "common-sense" findings (see Pettigrew, 1988). Perhaps, as a consequence, social psychology has evolved to distinguish itself with its clever methodologies and counterintuitive findings. As Hales (in press) discusses, Abelson, a leading and influential social psychologist at the time, broadly defined "interesting" findings as results that "change what scientists believe about important causal relationships" (Abelson, 1995, p. 158). Hales further asserts that to be "interesting," findings today

> must clear a much higher hurdle; they must be counterintuitive, yet still be grounded in predictions that could have been drawn from the existing literature. These features are not always compatible. In fact, they are often in direct competition. This creates pressure to engage in questionable research practices such as using researcher degrees of freedom (Simmons et al., 2011), or Hypothesizing After Results are Known (HARKing).

Jussim et al., in their article in the Special Issue, discuss a similar issue in terms of the "Wow Effect" (Jussim & Maoz, 2014) and the incentive to produce a "novel result that is seen as having far-reaching theoretical, methodological, or practical implications." Indeed, perhaps not coincidentally, the Open Science Collaboration (2015) reports that greater surprisingness of an effect is associated with greater failure in replication attempts.

Too much emphasis on new and impactful discovery in social psychology can lead to misguided research and publication standards, including insufficient attention to the replicability and stability of published (or publishable) findings. For example, in response to a concern articulated in an earlier crisis facing social psychology – questions about the relevance of the subdiscipline – finding significance with small samples was highly valued because it implicated large effect sizes (Bakan, 1966). Moreover, as Schaller (in press) describes in his article in this Special Issue, there is evidence that researchers tend to overestimate effect sizes associated with their findings and that, in general, these effects have much less impact in naturalistic settings than what the findings in the laboratory would suggest.

The crisis in replicability and the current threat to social psychology's scientific reputation and standing have painfully but commendably stimulated reflection upon the goals and values of social psychological science. It offers a rare opportunity to envision fundamental changes, rather than incremental ones. The theme of this Special Issue, while focused specifically on the topic of replicability, requires scholars to wrestle with the tension between (a) ensuring and documenting the stability of social psychological findings and (b) facilitating creativity and encouraging the rapid advancement of knowledge. In the remainder of this article, I propose that the field acknowledge and *embrace* the tension between the values of rapid advancement of knowledge and of replicability and stability of findings. Drawing regularly from the insights and arguments presented in the articles included in the Special Issue, I first comment on ways to ensure the stability and robustness of published findings in social psychology and then discuss the challenges of fostering creative, often counterintuitive, discovery – a hallmark of social psychology – in a scientifically responsible way.

2. Replicability and stability of published findings

There is currently broad consensus among social psychologists that replicability should be a central value in social psychology; it is also probably a necessary value, both politically and philosophically, for social psychology to be respected as a scientific discipline. Consistent with the theme of this Special Issue, several of the articles speak directly to replicability, but they address it at different levels. A number of articles present practical ways to enhance the replicability of particular types of research findings in social psychology. For example, Pirlott and MacKinnon (in press) recommend adopting more rigorous ways of establishing the causal role of proposed mediators in social psychological research. Other sets of authors raise broader questions about the most appropriate way to evaluate the replicability of the effect. Both types of approaches, specific and broad, are important to consider and debate. I highlight two particularly timely issues, one specific (electronic participant pools) and the other broad (replication assessment).

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

2.1. Replicability and electronic participant pools

Articles on this topic are particularly valuable because in social psychology, in which theory is prized, changes in more mundane aspects of methodology and their important implications are often overlooked. The articles by Berinsky, Margolis, and Sances (in press) and by Curran (in press) alert the field to potential challenges to the validity of social psychological findings based on increasingly prevalent online participant samples and describe ways to help ensure the validity, and thus replicability, of the research. The issues these articles raise also help illustrate an incentive problem that contributes to the replicability question that is the focus of the Special Issue. Researchers are highly incentivized to develop new ways to be productive in scholarship, while the publication system is structured in ways such that the impact of these innovations cannot be recognized until much later, after they have been regularly used and yielded a body of published research.

One of the historical criticisms of social psychology is that it is a science built on the responses of limited, nonrepresentative samples — specifically first- and second-year college students in introductory psychology courses. In part because of concerns about the nonrepresentativeness of college student participant pools but probably due to the convenience and the resource efficiency of new communication technology, social psychological science is being built increasingly on data using opt-in electronic participant pools (e.g., MTurk; see Buhrmester, Kwang, & Gosling, 2011). While the demographic range of participants in opt-in electronic pools is broader educationally, in age, and often socioeconomically than introductory psychology pools, these types of sample are not representative samples, and often – like other convenience samples – yield results different than what a representative sample would produce (Chang & Krosnick, 2009).

Articles in this Special Issue further emphasize that use of these pools requires new methodological and analytic standards. In particular, Berinsky et al. address the challenge of inattentive responders in these samples to the validity of research findings. As Berinsky et al. note, inattentive responders can be identified by embedded questions. One form of these questions resembles traditional manipulation checks, asking participants to recall key information related to the cover story and/or about the experimental manipulation. However, to accommodate the various contexts in which individuals from electronic participant pools respond (e.g., on work breaks, at home), "Screener" questions are often embedded in surveys or experimental questionnaires that, for example, ask participants their favorite color but also instruct them how to answer the question if they are being attentive.

While the social psychology laboratory could be criticized because it was not a naturalistic environment, it was a well-controlled environment. Thus, normally very few participants from the entire sample of participants were excluded from analyses, and mainly so because of suspicion or discerning the experimental hypothesis, rarely because of inattentiveness. Excluding participants because of responses they make has long been recognized as a threat to the internal validity because it undermines randomization (e.g., through subject selection or participant mortality; Campbell, 1957), and thus reduces experimental designs to correlational studies. In experiments that use electronic samples, it is not unusual to have 10%–20% of the original respondents excluded because of inattentiveness and the consequent "noise" they bring to the data. Curran, in his article in this Special Issue, argues that if participants who are careless or insufficiently attentive (CIE) responders are included in analyses, they

> pose a great threat to replicability in a number of ways. First and foremost, studies that inadvertently draw a larger proportion of these responders have a greater potential for false positives or false negatives. This may cause true findings to be hidden by these C/IE responses, and cause otherwise successful studies to languish in the file drawer. This may also cause spurious results to be published that are only a product of the high rate of C/IE responders in that sample, and not true underlying effects.

In addition, if exclusion of participants is systematically related to a manipulation, then internal validity – the ability to make causal inferences – is compromised. Moreover, as Berinsky et al. explain, even if there is no relationship between condition and excluded participants,

> these systematic differences present a problem for both external and internal validity. First, by excluding inattentive respondents — who are disproportionately younger, less educated, and non-white — researchers lose one of the main benefits of collecting data online, namely the ability to collect a diverse sample. Second, if attention is correlated with characteristics that interact with the treatment, the culled sample could produce an artificially large (or small) treatment effect.

Curran summarizes a range of techniques for detecting careless or inattentive responders and summarizes, in Table 1 of his article, offering various degrees of endorsement and ultimately recommending a multiple hurdles approach. In addition, Berinsky et al. identify specific strategies that can increase the number of participants who pass Screener items, but does not achieve the ultimate goal of reducing noise in the data. They conclude, "In sum, although we were able to successfully train and encourage respondents to pass Screener questions, these strategies do not improve the quality of our data; we were unable to turn shirkers into workers…. Our failure to convert shirkers in this manner strongly suggests that attentiveness to surveys is outside the realm of factors that researchers can control."

One of the most important contributions of the articles in this Special Issue is to alert the field not only to specific methodological challenges associated with current research practices, as the articles by Berinsky et al. and by Curran do, but also sensitize the field to a broader and recurring problem: Methodological innovation often occurs at a rapid rate, while the accumulation of data, recognition of potentially problematic implications, and publication of articles documenting a problem take much longer. Thus, widespread understanding of the validity of techniques typically occurs long after a body of work using those techniques has accumulated in the published literature.

2.2. Replicability and robustness of findings

Beyond the proximate goal of refining existing techniques or developing new ones to increase the internal validity of social psychological research, several articles in the Special Issue consider ways to assess whether specific findings are replicable. While there are points of overlap in the authors' discussions, there are also places of divergence in the perspectives and conclusions. Points of contention involve whether exact or direct replications should represent the standard for evaluating the stability and internal validity of social psychological findings, and whether it would be advisable to require findings be replicated by *other* researchers *before* the acceptance of manuscripts for publication.

With respect to the latter point, in his article in this Special Issue, Uhlmann (in press) proposes a process of pre-publication independent replication. He illustrates this process with the Pipeline Project in which 25 laboratories internationally engaged in direct replications of 10 moral judgment effects. He argues that direct replication attempts are not necessarily adversarial and ideally should be conducted before research is published. He explains that the process of pre-publication independent replication (PPIR)

> can help scholars ensure the validity and reproducibility of their emerging research streams. The benefit of PPIR is perhaps clearest in the case of "hot" new findings celebrated at conferences and likely headed toward publication in high-profile journals and widespread media coverage. In these cases, there is enormous benefit to ensuring the reliability of the work before it becomes common knowledge

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

among the general public. Correcting unreliable, but widely disseminated, findings post-publication.

The illustrative case of the replication project described by Uhlmann produced mixed results: Six findings replicated well, two with some qualification, and two findings did not replicate with the criterion that was set. Although there is evidence of considerable robustness of the effects examined, there was also more than enough nonreplication to be a cause for concern.

While recognizing these merits and the value to the field of avoiding "the cluttering of journal archives by original articles and replication reports that are separated in time, and even publication outlets, without any formal links to one another" (Uhlmann, in press), there are practical limitations to this approach. One is that it directs considerable resources in the field to replicating other scholars' findings – something many might argue is a net benefit rather than a cost – but also it does so in ways that might exclude researchers who do not have sufficient resources to be involved in such projects or who are not sufficiently well networked to be included in these large projects. A substantial portion of social psychological researchers work in settings that are not research intensive institutions, and one of the strengths of social psychology is that it values ideas and theory, and thus can be inclusive of scholars whose material research resources are limited. Another limitation is that some replications require considerable instrumentation, preparation, and time, and in an academic world with a limited "tenure clock," time is a critical commodity for junior scholars.

A theme in several of the articles in the Special Issue – most central in the articles by Crandall and Sherman (in press), by Fabrigar and Wegener (in press), and by Stroebe (in press) – is that many in the field are currently placing too much emphasis on making *exact* replications the standard for establishing the validity of findings. The arguments, to which I am sympathetic, are that (a) in social psychology, it may not be possible to distinguish exact replications from conceptual replications; and (b) there may be additional scientific benefits in terms of advancing scientific knowledge of pursuing conceptual rather than exact replications. Exact replications employ the same procedures and operationalizations as the original study to see if the same results can be obtained; conceptual replication tests the same fundamental hypothesis in the original research but in a way that triangulates on the basic finding in the original study using different operationalizations of the independent and dependent variables and often using a different participant population. (See also, in this Special Issue, Hüffmeier, Mazaei, and Schultze's (in press) paper, which outlines new comprehensive typology of types of replication studies.)

The articles by Crandall and Sherman, Fabrigar and Wegener, and Stroebe all point out the difficulty of having exact replications be truly exact, because of time, external events, or shifts in the demographics of the participant pool they may have different meaning to participants in the replications study. Demonstration of the psychometric invariance of the measures – evidence that the meaning of the measures from the original study to the replication(s) have similar meaning to participants – is thus needed. In cross-cultural psychology, showing psychometric invariance is normally expected. In social psychology, researchers typically aspire to reveal general principles of human behavior, usually ignoring the geographical context that participants bring to the study. However, these geographic influences, which are typically relegated to a contributor to error variance, are also systematic influences. Moreover, Fabrigar and Wegener note, "specific operationalizations used in many experiments are not constructed with the intent of making the operationalization per se broadly applicable across samples and contexts. Instead, the operationalizations are constructed to be optimal within the population and context in which they will be used." Thus, even within the US, the same measures likely have different meanings in different research contexts. As these articles collectively suggest, then, the distinction between direct and conceptual replication is not easily made, it is prudent to assume that, unless otherwise determined, any replication is conceptual (see also Stroebe & Strack, 2014).

The second type of argument against relying on exact replications for determining the replicability of social psychological findings involves questions about the primary goals of exact and conceptual replication. Crandall and Sherman state that 'exact replication may be effective, in their terms, for "ferreting out error" but limited in advancing the science.' Crandall and Sherman observe:

> An emphasis on direct replications focuses a field's attention and resources toward upholding versus upending specific findings, rather than toward building a theoretical framework that is robust across specific operationalizations. Direct replications take a careful look at the status quo, but rather than provide and test an alternative, they must accept the choices of a prior scientist or lab. They can support the finding or they can question the finding, but they cannot replace the finding.

Conceptual replication tests the same fundamental hypothesis in the original research but in a way that triangulates on the basic finding in the original study using different operationalizations of the independent and dependent variables and often using a different participant population. Besides questioning whether exact replication is even possible, Crandall and Sherman propose that conceptual replication can demonstrate the basic stability of a conceptual relationship while offering opportunities to gain incremental knowledge through identification of moderating factors and boundary conditions. Maner (in press), in his article in this Special Issue, recommends greater emphasis on field research to ensure the replicability and argues that attempts to conceptually replicate prior findings under circumstances "that reflect normal variability in the world and diverge from the original study … [can] provide greater confidence in the robustness of the findings."

The case for encouraging conceptual replication is a persuasive one, but it is a compromise between establishing the stability of a previously documented effect and the acquisition of new knowledge. Conceptual replications, which represent a "drilling down" approach, can advance knowledge systematically by identifying boundary conditions, moderators, and mediators for a particular phenomenon. But, because conceptual replications begin with the same base finding, the opportunity for *dis*continuous discovery, which can advance knowledge even more rapidly, is limited relative to research that has more conceptual degrees of freedom to adopt different assumptions and new perspectives.

The articles by Fabrigar and Wegener and by Stroebe further argue that the stability of a particular psychological finding might best be determined through meta-analytic techniques (but see the article by Sakaluk, in press, in this Special Issue). Meta-analytic procedures determine an overall significance level and effect size across a defined set of studies. While the standards for evaluating whether a finding is statistically significant ($p < .05$) and the interpretation of effect size are currently consensually established in the field, the standards for determining successful exact replications (e.g., 50%, 75%, 100% of independent replications that yield the original findings) have yet to be established. Given the arguments of the multiple factors that can reasonably be expected conceptually to vary across laboratories attempting to conduct exact replications, 100% rates of successful replication are unlikely. In that context, is the finding by Uhlmann that of the 10 findings tested, six replicated fully good news or bad news?

Meta-analysis can also assess potential heterogeneity of results, which can be used to suggest the existence of moderating influences. Within these analyses, a particular failure to replicate at the traditional .05 level of significance a previously observed finding is not by itself a cause for alarm or suspicion. It is a single data-point in an array of findings, from which meta-analysis can determine a general effect. As Stroebe explains in his article in this Special Issue, social psychology experiments – perhaps because of social psychology researchers' greater interest in effect sizes than the stability of an effect – have often been

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

underpowered. As a consequence, even a nonsignificant effect may contribute positively to the detection of the effect in a meta-analysis. Stroebe observes, "It is intriguing, that meta-analyses present a picture of the replicability of social psychological research that differs considerably from that suggested by mass replication." Social psychological findings appear more stable from the perspective of meta-analytic projects than they do from the results of replication projects. In the Pipeline Project described by Uhlmann, meta-analyzed effects for all 10 studies analyzed were in the predicted direction and showed corresponding significance in eight out of the 10 cases.

Fabrigar and Wegener also propose a particular application for meta-analysis for replications:

> once an effect has been initially demonstrated, the results of each new attempt to replicate this effect (presuming a credible case exists for assuming psychometric invariance) should be integrated with the original demonstrations and earlier replication attempts in a "cumulative meta-analysis" that provides an updated assessment of the overall statistical evidence for the effect.

Taken together, the articles in this Special Issue about the best way to assess the replicability of findings in social psychology lead to the kind of conclusion that develops when highly intelligent and informed people debate. There is considerable wisdom in each article, and the best solution involves multiply-operationalizing successful replication. That is, projects that focus on auditing the field through systematic attempts at replication as close to exact as possible are valuable for assessing the replicability and the stability of specific findings. However, because exact replication is primarily useful for detecting errors, they should be supplemented with tests of cumulative findings, including conceptual replications and extensions, conducted well after an original finding has been published.

Certainly, this is a delicate time in terms of social psychology's scientific standing, and the subdiscipline needs to be cautious about erroneous claims. Nevertheless, the scientific enterprise involves successive approximations to knowledge (Campbell, 1957) and should be self-correcting as research findings are accumulated. Moreover, given the changing nature of the social phenomena, which are affected not only by changes in political and social events but also by the dissemination of social psychological knowledge to the public (Gergen, 1973), failures to replicate an original finding at a later time may not mean the original finding was in error (that is, not internally valid) but that additional moderating elements may systematically affect the generalizability of the finding (that is, limit its external validity).

From this perspective, inability to replicate is not necessarily a failure in the science of social psychology. However, knowing whether a finding is replicable and enduring still is essential to the health of psychology as a science. Thus, incentivizing both replication projects and meta-analytic tests within the publication system is critical to the future of the social psychology. As Hales (in press) observes, "the concerns about replicability are intertwined with current research and publication practices, and that to solve the current crisis we will need to change the way we conduct and communicate research." In the next section, I again drawing heavily on insights, arguments, and recommendations from the articles in this Special Issue to discuss selected issues implicating the critical role of the publication and professional context in replication issues and more generally in the conduct of research.

3. Facilitating creativity and the advancement of knowledge

More regular and rigorous attention to determining whether an effect is replicable will not, by itself, ensure the validity of the body of social psychological literature. Researchers are human and are susceptible to the biases in perceptions and cognitive processes that social psychologists regularly study. Thus, as a consequence of their motivation to publish important findings, they will not, and perhaps in the long-run cannot, be dispassionate in the way they analyze and interpret their findings. Hales comments that these "practices are not used to actively deceive an audience. The researcher does not know whether their hypothesis is true or not, and is therefore not lying when engaging in these practices. However, these practices do prioritize finding significant results over finding results that accurately describe reality."

Several of the articles in this Special Issue thus propose that the field needs to institute policies that help distance social psychologists from their hypotheses. Indeed, people generally feel a strong sense of ownership over their ideas and closely associate them with their self-concept (Pierce, Kostova, & Dirks, 2003). Moreover, the public presentation of ideas, as in manuscripts submitted for publication, increases people's beliefs in the veracity of their statements (Echterhoff, Higgins, & Groll, 2005). In a well-meaning attempt to limit the passive voice, long condemned in writing-style books, current APA publication guidelines encourage authors to take ownership over their hypotheses (e.g., "We hypothesized" rather than "It was hypothesized").

Jussim et al. (in press) highlight how a range of biases exhibited by individual researchers and observable within the publication system can influence research interpretations, and thus the validity of conclusions. One example of this is the confirmation bias, a general type of motivated reasoning that occurs when people seek out and evaluate information in ways that support pre-existing views while discrediting information that opposes these positions. The confirmation bias systematically can distort scientific conclusions. As Jussim et al. explain, biases such as this "influence each phase of the research process, including how people interpret research findings." Jussim et al. also list a series of specific recommendations for facilitating awareness of potential bias and correcting for it in the review process.

Greenwald's (1975) survey of scholars in social psychology directly implicates the role of the confirmation bias in the scientific process. Greenwald found that, on average, researchers believed that their hypotheses were 83% likely to prove to be correct. Only 25% of researchers reported they would abandon their hypothesis after initial nonsignificant findings, whereas 60–70% of researchers and reviewers said they would pursue an exact or modified replication instead. About 60% of the researchers surveyed indicated that they would immediately pursue publication after they found statistically significant support for their hypothesis in an initial study. Thus, researchers attempt to confirm their hypotheses in the face of null results and require only limited supporting evidence to publish results consistent with their predictions. This is a recipe for inflating the number of false-positives in the literature.

To help researchers attain greater objectivity, Schaller (in press) recommends that hypotheses be framed as if-then propositions. He writes:

> Rather than tacitly assuming that "hypotheses" are nothing more than subjectively plausible personal predictions, hypotheses can instead be explicitly articulated as depersonalized statements that follow from the systematic application of "if-then" logic. The basic principles are simple, and intuitively appreciated by most psychological scientists: Some set of underlying assumptions or assertions are specified; and then some set of further implications (e.g., in the form of "if-then" statements) are logically derived. These logically derived implications have the logical status of hypotheses.

Adopting a different approach, Hales contends that drawing attention to the kinds of statistical errors that can be made, beyond Type I and Type II errors, can sensitize researchers to the complexity of statistical decision-making and help "researchers and evaluators of research draw more reliable conclusions." He defines Type III errors as instances in which researchers draw conclusions from poor evidence; Type IV errors involve the "incorrect interpretation of a correctly rejected hypothesis (Marascuilo & Levin, 1970 p. 398)."

These types of interventions and controls may help protect the field to some degree from invalid research findings and misrepresentations of results. Nevertheless, such efforts may be limited in their ultimate

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

success by these prevalent, almost unavoidable, human biases in terms of commitment to ideas and motivations to confirm them. In addition, the incentives of publication – including employment security and advancement, acceptance by and status among peers, and public attention and prestige – fuel powerful motivations to report novel and important findings.

Changes along these recommended lines can improve the conduct of research. However, the impact of these proposed policies may be restricted by limitations in imagination among reviewers and editors – the human factor – who have been socialized to give priority to "interesting" findings, in terms of results that change what research believe (Abelson, 1995). It is not uncommon (at least in my personal experience) to have a manuscript rejected because the conclusion of those involved in the editorial process was that, although the methodology was solid, the results were "intuitive and unsurprising."

An alternative approach would be structural. Leading journals could be re-structured to encourage both creative publication *and* replication (through multiple studies with a paper or through independent replication). This might be accomplished by having separate sections of the same journal (as some journals now distinguish research reports from research articles) dedicated to the different primary scientific objectives of (a) rapidly advancing knowledge (e.g., in a section titled, "New Directions," "Tentative Findings," and/or "Exploratory Findings"; *and* (b) establishing the replicability, stability, and robustness of particular findings (e.g., in a section titled, "Replication and Extension").

The Replication and Extension section might not only report exact replications but also include conceptual replications with the goal of theory development (see Hüffmeier et al.'s article). This section could also include, and thus promote, field studies that test the robustness of experimental findings, as suggested by Maner (in press). It might also be open to the kind of empirically-driven work that Greenwald, Pratkanis, Leippe, and Baumgardner (1986) proposed the field should encourage to complement theory-driven research.

The first step in this proposal to restructure major journals would be to define the values that socially psychologists believe are most important to the science of psychology, and then organize the journals in clear ways that facilitate research of this type. An advantage of classifying different types of research in different sections is that it can help achieve "truth in advertising." One section seeds new perspectives and scientific innovation but with the caution of caveat emptor (buyer beware), while the other section evaluates how well findings hold up. Thus, rather than wrestling with the tension between scientific certainty and creativity *within* every study, these goals may be able to coexist in a complementary way at a systems level. The virtue of having these different sections under the umbrella of same journal is that articles in the different sections all benefit in a similar way from the journal's reputation and prestige. Replication research and applied research can therefore gain the stature of "pure" research with "interesting" results.

I recognize that the proposal I presented can be criticized on a number of different grounds. On a practical level, decisions will have to be made by authors and editors about how a particular manuscript should be classified. But those are issues that occur in other guises. Is work more appropriate for a brief report or a longer research article? Should a manuscript on prejudice be submitted to a journal section on attitudes or on intergroup relations? In terms of substance, the journal restructuring idea may be viewed as too radical, because it "elevates" certain types of research to the level that has been reserved for the kinds or research that changes what scholars believe. In their article in this Special Issue, Hüffmeier et al. document current barriers to conducting and publishing replication research. For instance, journals that focus on applying social psychology or replications in social psychology have struggled to survive or achieve high levels of prestige (Giner-Sorolla, 2012). Alternatively, my proposal may be seen as far too modest, because it does not require every article to demonstrate the replicability of its effects. Reverse engineering, however, is a proven tool for accomplishing significant system change.

4. Conclusion

In conclusion, the articles in this Special Issue make important contributions in refining the field's understanding of the replicability problem in social psychology and helping to solve it. This commentary self-consciously built upon many of the specific proposals and views expressed in the individual articles and identified basic issues that tie these works together. In this commentary, I propose that although it is valuable to assess and enhance the replicability of research, to evaluate the impact of current methodologies, and to present modified or new techniques to improve the validity of the research process among individual investigators, these efforts are best considered within a systemic framework.

Specifically, I propose that, while both admirable and complementary in principle, the objectives of demonstrating the stability of results and the rapid advancement of new knowledge create tension in practice. The incentives of the traditional publication system place greater emphasis on novel findings than on replication, creating an imbalance of priorities for individual researchers. This prioritization has been at the root not only of the present crisis in social psychology but also has played a role in previous crises. The priority given to new, "interesting," and often counterintuitive findings in social psychology – a historically distinguishing characteristic of the field – encourages (through its impact on employment and recognition) many of problems the field now faces: replicability, p-hacking, and data falsification.

While addressing the replicability problem can be done incrementally, through many of the excellent insights offered by the authors in this Special Issue, the fundamental change that is needed is in the reward structure of the profession, and therefore in the publication system that is the gatekeeper for scholarly success. If the publication system formulates ways to comparably value creativity and rapid development of new knowledge, on the one hand, and evidence of the replicability, robustness, and external (as well as internal) validity of findings, on the other, it will produce the balance needed for a healthy science.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423–437. http://dx.doi.org/10.1037/h0020412.

Berinsky, A., Margolis, M., & Sances, M. (2016). Can we turn shirkers into workers? *Journal of Experimental Social Psychology* (in press).

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5. http://dx.doi.org/10.1177/1745691610393980.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312. http://dx.doi.org/10.1037/h0040950.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet. Comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*, 641–678. http://dx.doi.org/10.1093/poq/nfp075.

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications. *Journal of Experimental Social Psychology* (in press).

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology* in press.

Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology*, *89*, 257–276. http://dx.doi.org/10.1037/0022-3514.89.3.257.

Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology* (in press).

Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, *19*, 146–151. http://dx.doi.org/10.3658/s13423-012-0227-9.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*, 209–320. http://dx.doi.org/10.1037/h0034436.

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. http://dx.doi.org/10.1177/1745691612457576.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1–20. http://dx.doi.org/10.1037/h0076157.

Greenwald, A. G., Pratkanis, A. G., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, *93*, 216–229. http://dx.doi.org/10.1037/0033-295X.93.2.216.

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008

Hales, A. (2016). Does the conclusion follow from the evidence? Recommendations for improving research. *Journal of Experimental Social Psychology* (in press).

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3), e1002106. http://dx.doi.org/10.1371/journal.pbio.1002106.

Huffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology* (in press).

Jussim, L., & Maoz, I. (2014). Desperately seeking the "wow effect": Data interpretation and scientific story-telling as issues of research integrity. *Presentation at the October 2014 meeting of the society for experimental social psychology*.

Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology* (in press).

Maner, J. K. (2016). Into the wild: Field studies can increase both replicability and real-world impact. *Journal of Experimental Social Psychology* (in press).

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of type IV errors. *American Educational Research Journal*, *7*, 397–421 doi: http://www.jstor.org/stable/1161635.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*, 217–243. http://dx.doi.org/10.1080/1047840X.2012.692215.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631. http://dx.doi.org/10.1177/1745691612459058.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*. http://dx.doi.org/10.1126/science.aac4716 (aac4716).

Pettigrew, T. F. (1988). Influencing policy with social psychology. *Journal of Social Issues*, *44*(2), 205–219. http://dx.doi.org/10.1111/j.1540-4560.1988.tb02071.x.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, *7*, 84–107. http://dx.doi.org/10.1037/1089-2680.7.1.84.

Pirlott, A., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology* (in press).

Reis, H. T., & Stiller, J. (1992). Publication trends in JPSP: A three-decade review. *Personality and Social Psychology Bulletin*, *18*, 465–472. http://dx.doi.org/10.1177/0146167292184011.

Rodin, J. (1985). The application of social psychology. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology (3rd ed.)2*. (pp. 805–881). New York, NY: Random House.

Sakaluk, J.,. K. (2016). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology* (in press).

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* (in press).

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multi-study articles. *Psychological Methods*, *17*, 551–566. http://dx.doi.org/10.1037/a0029487.

Schultz, D. M. (2010). Rejection rates for journals publishing in the atmospheric sciences. *Bulletin of the American Meteorological Society*, *91*, 231–243. http://dx.doi.org/10.1175/2009BAMS2908.1.

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. http://dx.doi.org/10.1177/0956797611417632.

Stroebe, W. (2016). Are most published social psychological findings false? *Journal of Experimental Social Psychology* (in press).

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *8*, 59–71. http://dx.doi.org/10.1177/1745691613514450.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670–688. http://dx.doi.org/10.1177/1745691612460687.

Uhlmann, E. L. (2016). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *Journal of Experimental Social Psychology* (in press).

Please cite this article as: Dovidio, J.F., Commentary: A big problem requires a foundational change, *Journal of Experimental Social Psychology* (2016), http://dx.doi.org/10.1016/j.jesp.2016.01.008