# Solutions to the Credibility Crisis in Management Science

ELIZA K. BYINGTON
The University of Sydney, Sydney, Australia

WILL FELPS University of New South Wales, Sydney, Australia

We argue that much academic misconduct can be explained as the result of social dilemmas occurring at two levels of management science. First, the career benefits associated with engaging in noncredible research practices (NCRPs; e.g., data manipulation, fabricating results, data hoarding, undisclosed HARKing) result in many academics choosing self-interest over collective welfare. These perverse incentives derive from journal gatekeepers who are pressed into a similar social dilemma; namely, an individual journal's status (i.e., its "impact factor") is likely to suffer from unilaterally implementing practices that help ensure the credibility of management science claims (e.g., dedicating journal space to strict replications, crowd-sourcing replications, data-submission requirements, in-house analysis checks, registered reports, Open Practice badges). Fortunately, research on social dilemmas and collective action offers solutions. For example, journal editors could pledge to publish a certain number of credibility boosting articles contingent on a proportion of their "peer" journals doing the same. Details for successful implementation of conditional pledges, other social dilemma solutions—including actions for management academics who support changes in journal practices (e.g., reviewer boycotts/buycotts), and insights on credibilitysupportive journal practices from other fields—are provided.

At this moment, the scientific endeavor is at an important crossroads. On the one hand, the reach of scientific ideas and their potential to do good has never been greater. The availability of vast sources of electronic data, and exciting advances in analytic techniques are creating new possibilities for discovery. TED talks, podcasts, and popularpress books are fueling intense public interest in science—particularly in applied sciences such as management. And the Internet, Google Scholar, and the Open Access movement are making scholarship that can inform practice more available than ever before. However, at the same time that scientific claims are more available, a growing chorus of voices from across the sciences (including management) are raising concerns about whether these claims can be trusted as a basis for future research and practice (Bedeian, Taylor, & Miller, 2010; Bettis,

Authors appear in alphabetical order. Authors contributed equally.

2012; Bosco, Aguinis, Field, Pierce, & Dalton, 2015; Francis, 2014; Harrison, Banks, Pollack, O'Boyle, & Short, 2017; Honig, Lampel, Siegel, & Drnevich, 2014; Ioannidis, 2005; Kepes & McDaniel, 2013; Leung, 2011; Mezias & Regnier, 2007; Open Science Collaboration, 2015; Scandura & Williams, 2000; Simmons, Nelson, & Simonsohn, 2011; Singh, Ang, & Leong, 2003; Webster & Starbuck, 1988).

Although management scholarship undoubtedly contains useful truths, there is growing evidence that management researchers often engage in research practices that can produce false claims. In particular, in a recent survey of management scholars by Bedeian, Taylor, and Miller (2010), the vast majority of respondents report having recently heard about or directly observed scholars engaging in questionable research practices.

These practices may explain why studies that attempt to reproduce management claims fail far more frequently than they should—with failure to

completely reproduce prior results occurring over half the time (Hubbard & Vetter, 1996; see also Open Science Collaboration, 2015), 10 times more frequently than one would expect based on standard levels of statistical significance (p < .05). Originally strong results are often either weaker or nonexistent when future scholars attempt to reproduce them (Webster & Starbuck, 1988). Such findings have led journalists to wonder if there is something wrong with the scientific method, posing the question: "Does the truth wear off?" (Lehrer, 2010).

The effects of these practices also appear to be filtering up to the metalevel, with a flurry of recent papers calling management's meta-analytic results into question, claiming that many meta-analyses may be reporting inflated effect sizes that do not accurately capture the true underlying relationships. This dynamic has been observed in strategic management research (Harrison et al., 2014), entrepreneurship research (O'Boyle, Rutherford, & Banks, 2014), judgment- and decision-making research (Renkewitz, Fuchs, & Fiedler, 2011), business ethics (Orlitzky, 2011), and industrial-organizational psychology (Kepes & McDaniel, 2013). As such, even meta-analytic management conclusions may be suspect.

Finally, management science is attracting concerns about the credibility of its claims from both internal and external stakeholders. Management has recently suffered a high-profile scandal involving data integrity and manipulation (Matlack, 2013). Business journals have been singled out as being slow to address credibility issues (Karabag & Berggren, 2012; Marcus & Oransky, 2014). And other scientists are raising concerns about the credibility of statistical results reported in individual management articles on PubPeer (PubPeer, 2015). Indeed, management claims may be particularly vulnerable to such critiques because, as Hambrick (2007: 1350) notes, "[T]he vast majority of published ideas in management are submitted to no tests at all, a handful are submitted to one test, and only a minuscule few are tested over and over or in multiple ways. Again, we don't really know much for sure."

Evidence of these problems in management has been steadily accruing for decades (e.g., Hubbard & Vetter, 1996; Webster & Starbuck, 1988). And yet, little has changed, despite appeals for scholars and journals to change their practices to produce more credible science (Bedeian et al., 2010; Bettis, 2012; Honig et al., 2014; Kepes, Bennett, & McDaniel, 2014;

Kepes & McDaniel, 2013; Leung, 2011; Mezias & Regnier, 2007; Pfeffer, 2007).

Our article offers (1) a new diagnosis of the drivers of noncredible science, researcher misconduct, and the lack of change, as well as (2) a new set of pragmatic prescriptions to help address these issues. In offering these solutions, we aim to provide options that recognize the mixed motives of most actors for both self-interest and collective welfare. Our analysis suggests that the credibility crisis will not be resolved by scholars' self-policing or unilateral changes in a particular journal's policies.

To preview the structure of this article, we follow a causal chain that works backward to identify the root cause of noncredible claims in management. In particular, we first synthesize arguments that noncredible research practices (NCRPs) on the part of individual management scholars can produce both noncredible conclusions and career benefits for individual researchers. As a result, we conclude that management scholars are trapped in an iterated public-goods social dilemma, where the actions that benefit individual scholars are bad for the field as a whole. Second, we argue that NCRPs could be reduced by a variety of credibility-supportive journal practices (CSJPs)—particularly new approaches that have emerged in other scientific fields (e.g., Open Practice badges for articles with credibility-enhancing features, accepting articles for publication "results-blind," etc.). The adoption of these credibility-supportive journal practices could reduce or remove the social dilemma for scholars by simultaneously rewarding credible contributions and reducing the benefits of noncredible research practices. Third, we further extend the current discussion of the credibility crisis in management by noting that journals are unlikely to adopt such policies. Working backward toward root causes, we suggest that the lack of widespread adoption of credibility-supportive journal practices is at least partly a function journal gatekeepers'—editors and reviewers—desire to protect and enhance their journal's status. In particular, we argue that the unilateral adoption of credibility-supportive journal practices will diminish the focal journal's "impact factor." Thus, we suggest that journal gatekeepers are also caught in a social dilemma, such that protecting their journal's status is at odds with adopting CSJPs that could benefit the field as a whole. We argue that resolving this social dilemma requires collective action. This can be accomplished by leveraging insights from the social dilemma and social movement literatures (Balliet, 2010; Chen,

1996; Opp, 2009; Zeng & Chen, 2003). In particular, we discuss opportunities for addressing this social dilemma with structural solutions (i.e., conditional pledges, defining small peer-journal groups, and reviewer pledges) and motivational solutions (i.e., multi-journal communication, and injecting a moral frame). Figure 1 provides a roadmap for our arguments.

#### NONCREDIBLE RESEARCH PRACTICES

In this section, we briefly review the credibility-reducing practices that can lead to the noncredible results discussed above, and the career motives for engaging in them. Specifically, we consider four noncredible research practices: data manipulation (e.g., dropping data that undermines results, p-Hacking), outright fabrication (e.g., creating or altering raw data), refusing to share data with noncollaborators, and unreported HARKing.<sup>1</sup>

#### Data Manipulation

Depending on the research design, data manipulation can occur in multiple ways. For example, scholars can select only those data that support a hypothesis and withhold the rest. Researchers can drop "uncooperative" measures from analyses. Similarly, researchers can choose control variables and statistical techniques based on those that yield supportive results. How do such actions affect the credibility of scientific claims?

A simulation by Simmons, Nelson, and Simonsohn (2011) demonstrates that a study can produce false positives 60.7% of the time ( $p \le .05$ ) if scholars have the following four "researcher degrees of freedom:" (1) the choice between two dependent variables, (2) the possibility of collecting 10 more observations per cell, (3) a variable that can be used as either a control or moderating variable, and (4) the ability to drop one of three conditions (or not). Given that so many studies have at least these "degrees of freedom," it is quite possible for scholars to change study results without actually fabricating data (see also, Goldfarb & King, 2016). In sum, we predict:

Pla: Data manipulation leads to less credible claims.

We predict that data manipulation is beneficial for scholar's careers given the clear benefits of such manipulations for producing "highly publishable" findings, and the indirect evidence that this practice is common. Bedeian and colleagues' (2010) survey finds that withholding nonsupportive data are a practice that 77% of sampled management scholars reported observing or hearing about within the prior year, and 60% reported observing or hearing about dropping data points based on a "gut feeling" within the prior year. Simmons and colleagues note that such practices are likely not malicious, but rather result from scholars feeling pressure to find significant effects, and vague guidelines as to appropriate practices for collecting and processing data (2011). Thus, we propose:

P2a: Data manipulation is beneficial for management scholars' careers.

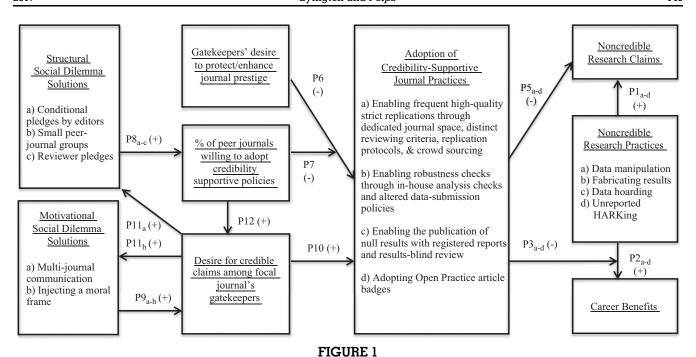
#### Fabricating Results

Of course, if scholars make up data or change data to support hypotheses, subsequent analyses are not credible. Incidents of wholesale and sustained fabrication of results are probably rarer than the other noncredible research practices (Fanelli, 2009). However, in the study by Bedeian and colleagues (2010), a sizable 27% reported that they knew of scholars fabricating results. Indeed, using methods that correct for scholar's reluctance to admit fabricating results, John, Loewenstein, and Prelec (2012) estimate that almost 40% of scholars may have done so at some point in their careers. This suggests that the temptation to change a few data points to make a hypothesis statistically significant is likely to occur in some published research and, like data manipulation, data fraud may be more common than generally realized.

P1b: Fabricating results leads to less credible claims.

Fabricated results can have severe negative career consequences if detected, as in the recent cases of Diederick Stapel (Bhattacharjee, 2013) and Ulrich Lichtenthaler (Matlack, 2013). However, because management scholars are rarely required to produce the data associated with their manuscripts, the likelihood of detecting fabrication seems to be

<sup>&</sup>lt;sup>1</sup> In this article, we focus on four Noncredible Research Practices that can undermine the credibility of research conclusions. These are not the only morally questionable research practices: see Bedeian and colleagues (2010). However, other ethically questionable practices (e.g., reciprocal authorship, plagiarism, and double publishing) are less related to the production of untrustworthy claims.



Causes of and Solutions to Management's Credibility Crisis. For the sake of simplicity, Propositions  $4\alpha$ -d have been omitted from the model. Propositions  $4\alpha$ -d suggest a feedback loop, where scholars will engage in Noncredible Research Practices less if those practices do not provide career benefits (i.e., if P3 $\alpha$ -d is true).

minimal. And indeed, in both cases mentioned above, concerns about the veracity of data were not raised in the course of the peer-review process (the primary instance of credibility checking research receives), but rather, on the initiative of concerned readers (Oransky, 2012) and a PhD student (Bhattacharjee, 2013). Until they were caught, these authors' manufactured findings led to speedy promotion, accretion of resources, and recognition as rising stars (Bhattacharjee, 2013; Handelsblatt Business Administration Ranking, 2009). As noted above, some estimates suggest that data tampering is fairly common (John et al., 2012). However, even conservative estimates (1-2% of scholars; Fanelli, 2009) would mean that hundreds of management scholars have fabricated results at some point, while only a handful of instances have been detected. This suggests that the vast majority of fabrication goes undetected. Thus, given few checks of data integrity:

P2b: In general, fabricating results is beneficial for management scholars' careers.

#### Data Hoarding

Sharing the data associated with reported scientific claims can help boost a publication's credibility in

two key ways: First, sharing original data allows for an independent analysis to check the accuracy of reported results (what have been called "analysis checks;" Tsang & Kwan, 1999). Given the potential for data manipulation and fabrication discussed above, along with the possibility of honest errors given the complexities of database management and analysis, sharing data boosts credibility by allowing for verification. Second, sharing data allows for what have been called "reanalysis checks" (Tsang & Kwan, 1999) where, for example, a different analytic technique is applied to the data associated with prior claims to assess the claim's robustness. This can occur when superior analytic techniques are developed and applied to prior data sets. In contrast, refusing to share the data on which claims are based—that is, "data hoarding"—prevents these credibility checks of claims from being performed.

Plc: Data hoarding leads to less credible claims.

Although sharing data enables the independent assessment of claims, Blanton and colleagues (2009a) found that half of scholars from whom original data were requested "stated that they either would not or could not comply with so basic

a scientific norm as making their data available for reanalysis" (Blanton et al., 2009b: 598). Similarly, an effort at reanalysis in social psychology found that authors of 73% of the articles in their sample (out of 249 articles) would not provide their data (Wicherts, Borsboom, Kats, & Molenaar, 2006). This is in spite of these specific scholars having signed an American Psychological Association statement promising to do so upon request.

Data hoarding by scholars may be due, at in least part, to career-related motivations. First, restricting access to a data set ensures that study authors control how data are used. For example, data hoarding allows original study authors to negotiate coauthorship in exchange for providing their data for a future study. Moreover, restricting access removes the risk that others will "scoop" the original study authors on their intended future research using their own data.

Second, data hoarding prevents other scholars from reanalyzing the original data in ways that call into question the original publication and could result in the authors' loss of career status (e.g., highlighting errors that require correction or retraction). This appears to be a strong motivator of whether scholars share their data, as, for example, Wicherts, Bakker, and Molenaar provide concerning evidence from psychology that "willingness to share data when requested" is associated with a "higher prevalence of apparent errors in the reporting of statistical results. The unwillingness to share data was particularly clear when reporting errors had a bearing on statistical significance" (2011: 1). Thus, the proposed motivation for much data hoarding is ultimately the same as the proposed motivation for the other noncredible research practices: the desire to have a strong publication record with its consequent career benefits.

P2c: Data hoarding is beneficial for management scholars' careers.

#### Undisclosed HARKing

One of the most common noncredible research practices is presenting a post hoc hypothesis as if it

were an a priori hypothesis (unreported HARKing, that is, Hypothesizing After Results are Known (Kerr, 1998). This practice is also known as "data fishing" or "data dredging." From the standpoint of producing credible claims, the main reason unreported HARKing is problematic is not that it involves deceptive reporting (although it does). Rather, undisclosed HARKing is problematic because it is inductive research posing as deductive research, and is, thereby, held to a weaker statistical threshold. As others have described (Kerr, 1998; Leung, 2011), whenever multiple potential relationships in the same "family" are tested, each test needs to meet a higher standard of significance to account for all the other tests (e.g., Holm-Boneronni corrections). When such corrections are not applied, the likelihood of reporting false discoveries (i.e., false positives or Type 1 errors) increases dramatically (Shaffer, 1995). Indeed, fields relying on inductive theory building (e.g., genomics) use a variety of exploratory analytic methods designed to take into account the multiple relationships tested (Benjamini, 2010). In contrast, when scholars apply weak statistical standards (e.g., p < .05) to HARKed results, it is quite possible to manufacture statistically significant effects when none actually exist (Goldfarb & King, 2016; Leung, 2011; Simmons et al., 2011). Thus, disclosed data fishing is not necessarily problematic as long as the standards applied reflect the number of hypotheses tested. However, if not disclosed, HARKing inflates true effect sizes substantially (Bosco et al., 2015). This is no trivial concern, as "HARKing can lead to lessthan-ideal management practices because effect size estimates are the central input to estimates of practical significance" (2015: 35).

Pld: Undisclosed HARKing leads to less credible claims.

Despite these negative effects, HARKing is common. Approximately 92% of management scholars surveyed by Bedeian, Taylor, and Miller (2010) report observing or hearing about colleagues engaging in HARKing within the prior year. Similarly, Bosco and colleagues' (2015) find that 38% of corresponding authors associated with a sample of articles from Personnel Psychology and Journal of Applied Psychology self-reported that "at least one hypothesis had changed between the completion of data collection and publication" (2015: 11). And yet, while Kerr's (1998) survey of social scientists suggests that some sort of HARKing occurs in most social science

<sup>&</sup>lt;sup>2</sup> In some cases, there are important reasons for restricting data access; e.g., agreements with raw data providers, confidentiality (when data are impossible to anonymize), trade secrets, and national security (OECD, 2007). We do not contend that data hoarding is always and only associated with self-interested motives. Rather, we suggest that data hoarding can be motivated by career interests, and as the evidence suggests, that willingness to share data can also be related to the credibility of research claims (i.e., Wicherts, Bakker, & Molenaar, 2011).

research, most respondents also report that they believe HARKing is harmful for science and should be practiced less. Given scholars' concerns and its negative effects for science, why does HARKing occur?

We join a variety of others in recognizing that HARKed findings can be beneficial for one's career (Bosco et al., 2015; Kerr, 1998; Nosek, Spies, & Motyl, 2012). In particular, engaging in unreported HARKing can increase the chance that a scholar can identify statistically significant relationships: an important criterion for publishing research in management journals (Bettis, 2012), which is beneficial for an individual scholar's career outcomes (e.g., promotion, tenure, research funding, etc.). In contrast, honestly reporting HARKing when a priori statistical methods and standards were used may result in editors or reviewers requiring that claims meet much more stringent statistical significance levels to account for all the relationships tested, potentially making reported results statistically nonsignificant.

P2d: Unreported HARKing is beneficial for management scholars' careers.

#### THE SCHOLAR'S DILEMMA

Given that noncredible research practices can benefit a scholar's career, we argue that scientists are in a social dilemma that has far-reaching consequences for the field. By rewarding the behaviors associated with manufacturing "significant" effects in empirical scholarship (data manipulation, fabricating results, data hoarding, unreported HARKing), such behaviors become consistent with rational self-interest (Briner & Walshe, 2013; Nosek et al., 2012; Schmidt & Landers, 2013), but harmful for the field as a whole.

Conversely, we might imagine a "purist" who refuses to publish results unless they are confident that the results are true, and as such, does not in engage in any of the NCRPs described above. It is our hypothesis that such "methodological purists" will have worse career outcomes than scholars engaging in noncredible research practices. In particular, we contend that methodological purists will be less likely to publish papers in prestigious journals, get jobs at research-supportive universities, be awarded tenure, acquire social prestige, or go on to select, promote, and train others (e.g., supervise PhD students). Further, scientific asceticism may not be an attractive quality for potential collaborators. These predictions are troubling because they suggest that noncredible research practices are likely to become predominant in the field over time.

To be analytically precise, our contention is that the field places scholars in an iterated public-goods social dilemma (Balliet, 2010; Dawes, 1980), where (a) each publication submission involves a choice between "cooperation" (i.e., contributing to community interests by engaging in credible scientific practices) and "defection" (i.e., engaging in noncredible research practices), (b) the rewards for defection outweigh the rewards for cooperation in any given choice opportunity, but all individuals are better off if all cooperate than if all defect (the definition of a social dilemma), (c) higher payoffs from past choices mean more opportunities to make future choices (e.g., "cooperation" reduces the likelihood of surviving the tenure process; "defecting" garners greater research support), and (d) there is some social learning, such that other actors' cooperation or defection strategies are observed and "high performers" are emulated.

Above, we have described the variety and prevalence of practices by scholars that undermine the credibility of scientific claims. In doing so, we also noted evidence that many scholars (1) are aware that these practices are bad for the credibility of management science, and further, (2) think such practices should be done less (Kerr, 1998). This suggests that the problem of noncredible practices may be a "don't hate the player, hate the game" scenario. If we accept that scholars are just as likely to be self-interested and opportunistic as other humans, then noncredible research will be common as long as it is rewarded. To expect otherwise is to reward A while hoping for B.

"If we accept that scholars are just as likely to be self-interested and opportunistic as other humans, then noncredible research will be common as long as it is rewarded. To expect otherwise is to reward A while hoping for B."

#### CREDIBILITY-SUPPORTIVE JOURNAL PRACTICES

Solving the scholar's social dilemma would seem to require that journals adopt policies that make noncredible research practices less rewarding and encourage credible practices and claims. Fortunately, management need not "reinvent the wheel." Scholars and journals can benefit from building on the best practices of other disciplines that are

developing creative strategies to improve the credibility of their claims. In what follows, we review a variety of such credibility-supportive journal practices, most of which originated in other fields. These practices include (1) enabling frequent, highquality, strict replications through dedicated journal space, distinct reviewing criteria, replication protocols, and crowd-sourcing, (2) enabling robustness checks through in-house analysis checks and data submission policies, (3) enabling the publication of null results through registered reports and results-blind review, and (4) adopting Open Practice badges to recognize articles with credibility boosting features. As detailed below, these practices incentivize research that boosts the credibility of a field's claims and decrease the value of work produced using NCRPs. We consider each in turn.

## Enabling Frequent, High-Quality Strict Replications

#### Importance of Strict Replications

After a spate of retractions, a recent statement from the editorial leadership at Leadership Quarterly points out that although there are different perspectives on science and its aims, "Ultimately, however, science is based on the potential replicability of findings" (Atwater, Mumford, Schriesheim, & Yammarino, 2014: 1176). The goal of a strict replication is to assess the reproducibility of a prior result, that is, overlapping confidence intervals in effect sizes between studies using the original approach (Hubbard & Lindsay, 2013). As such, strict replications attempt to approximate the parameters of the original study as closely as possible (e.g., the research context, sample, and methods), with what Berthon and colleagues describe as "zero degrees of freedom" (2002: 425).3 Although it is impossible to exactly replicate the original study (at a minimum, time will have passed), the degree of correspondence between the original study and a "strict" replication, as well as the number of strict replication tests offered, can speak to the reliability of the original claim's credibility as a basis for future

Strict replications serve the function of identifying if original results might have been due to (1) a chance result, (2) key elements missing from the

original study's method description, and/or (3) noncredible research practices (Schmidt, 2009). When strict replications do not support the original claim, it highlights that the original study's described methods and claims may not provide a reliable basis for future action. In contrast, Tsang and Kwan emphasize that a confirmatory replication provides a "quantum leap of credibility" (1999: 110).

It is important to note that not all replications are "strict replications." Other types of replications (i.e., convergent, constructive, quasi-, corroborative, improvisational, modified, imprecise, or conceptual) deliberately vary aspects of the original study, such as the measures used, type of analysis, or population. As such, these types of replications cannot speak to the reproducibility of the original claim or the strength of its methodology to observe the effect as reported, since differences from (or even affirmation of) original results may be due to the differences between studies. For this reason, Nosek, Spies, and Motyl point out that replications that do not attempt to closely adhere to the parameters of the original study are "not an effective replacement for direct [strict] replication" (2012: 619), because they do not provide evidence of the reproducibility of the original findings. As Kepes and McDaniel point out, "The lack of exact replication studies prevents the opportunity to disconfirm research results" (2013: 257)—particularly an issue when noncredible claims are present in a field, and when there may be a bias for publishing only statistically supported claims.

#### Increasing the Frequency of Strict Replications

Dedicating journal space. Although providing evidence of reproducibility is core to the scientific endeavor, few management articles claim to provide strict replications of prior scholarship. For example, we manually reviewed 1,107 abstracts of articles published in the Academy of Management Journal over a 15-year period from 1998 to 2013 and found only one that claimed to provide a strict replication (i.e., Beyer et al., 1997).

Increasing the frequency of strict replications could be accomplished by allocating, for example, 10% of journal space to strict replications (Hubbard, Vetter, & Little, 1998; Nord, 1985). Alternatively, a journal could sponsor a yearly "special issue" of strict replications. Devoting space to replications has begun in psychology (Association for Psychological Science, 2015a), and is occurring in a forthcoming special issue in *Strategic Management Journal (SMJ* editors, 2014), which could provide exemplars.

<sup>&</sup>lt;sup>3</sup> Very strict adherence to original study parameters is likely to be easier in research designs that provide greater control (e.g., laboratory settings, vignette studies, studies with student samples) than in field studies.

Developing reviewing criteria for strict replication studies. As noted by Hambrick, "Take it from someone who has tried ... reviewers will come right out and laugh at you if you claim to be replicating a prior test of a theory" (2007: 1350). Providing reviewers with a set of criteria for assessing manuscripts that offer strict replications can serve to increase their number in management journals by (a) helping legitimize reproducibility-assessing scholarship in the minds of editors and reviewers, and (b) recognizing that the goal of strict replication research differs from that of research devoted to advancing theory-adding claims, and thus, should be assessed on different criteria (e.g., degree of adherence to the original study, sample size, number of strict replication attempts provided, the importance of the original claim in the literature).

#### Enhancing the Quality of Strict Replications

Expanded methods sections. Brief methods sections that do not provide sufficient detail for subsequent scholars to strictly adhere to the parameters of the original study can curtail attempts to reproduce results. Indeed, having the space to articulate various study design elements is not only key for enabling others to assess the reproducibility of claims, it is also important for understanding what the original study actually found (Williams, 2015).

Recognizing this issue, guidelines from a consortium of journals (e.g., Nature, Science, Cell, and other signatories) in partnership with the National Institutes of Health, advocate that journals provide unlimited (or very generous) space for methods sections (Marcus, 2014). This policy can be implemented by providing authors unlimited space in online appendices that do not require printed journal pages. Making method details available can facilitate high-quality strict replications that identify nonreliable claims, reducing the incentive to engage in NCRPs (see below).

Replication protocols. The Open Science Framework recommends that authors provide a replication protocol as an addendum to their manuscript that clearly outlines how to reproduce their study. Replication protocols can be posted in an online appendix, thus eliminating "the need for brevity" as a reason for omitting these critical details. Submitting a replication protocol as part of the review process also allows reviewers to ensure that the methods in a submitted manuscript are fully articulated and appropriate. In addition, submitting a replication protocol gives authors the opportunity to contribute to the

quality of subsequent attempts to replicate their work, and makes salient that replications of their claims will likely be forthcoming. Finally, replication protocols can also be added retroactively to published articles. As such, requesting replication protocols from authors may help engender a shift in the culture of management science toward replication-oriented thinking (Hubbard, 2015).

Crowd-sourcing replications. A barrier to drawing strong inferences from strict replications is that they could be statistical flukes. To overcome this challenge, an interesting new credibilitysupportive journal practice in psychology is the "crowd-sourced replication" (Nosek et al., 2012; Open Science Collaboration, 2015). For example, Perspectives on Psychological Science facilitates multi-university strict replication attempts, where approved research design protocols are "posted publicly, and other laboratories can follow the same protocol in conducting their own replications of the original result" (Association for Psychological Science, 2015a). All scholars who meet the procedural requirements have their collective findings published as a highly credible, multisite assessment of claim replicability and effect size. As Nosek, Spies, and Motyl point out, "as an open project, many collaborators [can] join and make small contributions that accumulate into a large-scale investigation" (2012: 622). Crowd-sourcing replications is a comparatively "low-cost" way for scholars to participate in credibility boosting publications for the field, and offers an attractive effortoutcome ratio for contributing scholars.

Original study authors may also be invited to submit (peer reviewed) commentary on the collated results of these studies, creating a career incentive for sharing insights on important aspects of the original research. Early evidence suggests that such replication efforts are often well-received by original authors, can identify key aspects of the original research that were underspecified (Schmidt, 2009), and can spark enlightening discussions about phenomena that make important contributions to science in and of themselves (Gong et al., 2014).

## Effects of Increasing the Frequency of High-Quality Strict Replications

Increasing and enabling the publication of strict replications can decrease the incentive to engage in NCRPs because they tend to boost the status of the most credible claims (and those who consistently produce reproducible findings). Thus, more strict replications could weaken the positive relationship between NCRPs and career benefits. If there is a very good chance that one's research will be strictly replicated, we suspect that many scholars would take greater care to use credible methods that allow them to be very confident in their results.

P3a: Journal policies that enable the frequent publication of high-quality strict replications would reduce (i.e., negatively moderate) the otherwise positive relationship between noncredible research practices and career benefits.

If P3a holds, then self-interested scholars will engage less in NCRPs. Further, devoting journal space to strict replications, providing more complete methods sections and replication protocols, and creating open-access opportunities for scholars to participate in high-quality crowd-sourced replications incentivizes scholars to produce scholarship that contributes to the credibility of the field, likely increasing the comparative attractiveness of such scholarship. Thus:

P4a: Journal policies that enable the frequent publication of high-quality strict replications would decrease the prevalence of noncredible research practices.

P5a: Journal policies that enable the frequent publication of high-quality strict replications would decrease the prevalence of noncredible research claims.

Enabling robustness checks. Although strict replications involve collecting new data, some types of scholarship draw directly on the data or information from the original study to draw conclusions about whether the findings are robust. These "robustness checks" include analysis checks, reanalysis checks, and refutations. Each is considered in turn.

### In-House Analysis Checking and Data Submission As Part of the Review Process

One way journals can prevent noncredible claims is to perform in-house analysis checks as part of the review process. Analysis checks apply the same methods to study data to ensure there are no errors in reporting. Indeed, in the closely related field of psychology, published errors appear to be quite common (Bakker & Wicherts, 2011), and concerns over the accuracy of analyses presented in management articles have also been raised (Credé & Harms, 2015). To counteract such errors, NIH guidelines developed with journal editors requires that "journal[s] should have a mechanism to check the statistical accuracy of submissions" (National Institutes of Health, 2015). Analysis checks could be performed by a methodologically oriented journal editor (perhaps an appointed role) or an in-house statistician. This avoids printing erroneous claims, and minimizes the need for corrections and retractions. Ideally, checking analysis would occur before the review process, thereby preserving reviewer resources when data are clearly flawed.

A key policy change that would be necessary for in-house analysis checks to occur would be requiring authors to submit anonymized data along with a codebook in a standardized format as part of the manuscript-submission process (or upon acceptance of a manuscript for publication). For the purposes of checking analysis during the review process, access to data could be restricted to authorized individuals only (e.g., a journal's in-house statistician). This can protect authors from concerns about loss of intellectual property. If analysis checks were common, it would discourage the NRCPs of data manipulation and fabricating results, because this could be uncovered during analysis checking (Simonsohn, 2013).

## Enabling Reanalysis Checks Through the Use of Social Science Data Repositories, Journal Data-Submission Policies, and Data Embargos

Reanalysis checks involve an independent scholar (team) applying alternate measurement or analytic techniques to data from a prior publication. Usually, these reanalyses aim to provide more accurate or additional information about a phenomenon and can yield highly valuable new information or aid in the interpretation of the original claim (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). The few reanalyses that have been published in the organizational literature have revealed that alternative statistical specifications and techniques can lead to very different conclusions than those in the original study (Blanton et al., 2009a; Drazin & Kazanjian, 1990; Edwards, Scully, & Brtek, 2000; Jones, 1992). However, it is difficult to know whether more published reanalyses would ultimately be a source of revision or reinforcement of management claims, because such research is very rarely published. This absence is especially disappointing in areas where patently superior statistical techniques that could be used to reassess prior data have become available since the original publication.

Journals could facilitate reanalysis checks by requiring the permanent posting of data along with a variable codebook in an approved online data repository (Wicherts et al., 2006). Indeed, numerous journal-vetted, quantitative- and qualitative-data repositories (and exemplar submission policies) already exist (e.g., *Nature*, 2015; Pampel et al., 2013; re3data.org, 2014). This is a result of a growing movement for publicly funded research data to be made available (Sutton, 2013) and journal policies requiring data availability (Rasmussen & Blank, 2007).

Management can learn from other fields how to execute this credibility-supportive journal practice. For example, in economics, they have learned a great deal about how to ensure that data and codebooks are organized in ways that allow for reanalysis (McCullough, McGeary, & Harrison, 2006, 2008). Moreover, to avoid discouraging scholars from collecting original data, it is possible to "embargo" data (DRYAD, 2015). Embargoed data are held in a data repository, not made public for a prespecified amount of time (e.g., 3 years), giving original study authors a "head start" before their data are released.

By facilitating reanalysis, data repositories should reduce the incentive to use NRCPs. For example, producing tenuous relationships using NCRPs would be less rewarding if independent analysis is likely to show that the prior findings depend on an arbitrary or inappropriate methodological choice. As noted by Simonsohn, "If journals, granting agencies, universities, or other entities overseeing research promoted or required data posting, it seems inevitable that fraud would be reduced" (2013: 1875). Also important, supportive reanalyses could boost the status of original study results.

Having a data-submission policy at journals raises the question: "What can be done when authors invoke privacy agreements with data providers, or claim that even anonymized versions of their data are too sensitive to post in research data repositories due to confidentiality concerns?" The above-mentioned NIH guideline of in-house statistical analysis checks by journals provides one approach to this issue. Another route is to update journal instructions to authors to state that (a) authors must formally identify their data as unshareable upon manuscript submission, (b) manuscripts based on unshareable data will be seen as less credible/weaker contributions to management science in the review process, since conclusions

based on unshareable data are less credible (Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012), and (c) that data were not provided by authors will be noted in any resulting publications as a caveat emptor for readers and scholars who might build on the work. Taken in conjunction, these journal policies should encourage authors to make their data shareable if at all possible, without obviating the possibility of publishing claims based on truly confidential data.

#### Refutation Publication Policies

Refutations provide critiques of a previous publication's research methods or conclusions (e.g., Fox, 2010). Publication of such statements can help identify problems before they negatively impact future research and practice. One way to improve credibility would be for journals to publish any valid refutations of prior work published in their journal. For example, in collaboration with the National Institutes of Health, the editors of Nature, Science, and a variety of other journals have developed a set of standards for journal editor signatories that includes "[h]av[ing] a policy stating that if the journal publishes a paper, it assumes responsibility to consider publication of refutations of that paper, according to its usual standards of quality" (Marcus, 2014: 966). Publishing refutations can incentivize the identification of extant research that contains inaccuracies and noncredible claims. These public critiques reduce the incentive to engage in NCRPs. In sum:

P3b: Journal policies that enable robustness checks would negatively moderate the otherwise positive relationship between NCRPs and career benefits.

If P3b holds, then self-interested scholars will engage in NCRPs less. Thus:

P4b: Journal policies that enable robustness checks would decrease the prevalence of noncredible research practices.

P5b: Journal policies that enable robustness checks would decrease the prevalence of noncredible research claims.

#### Journal Practices That Enable the Publication of Null Results

Null findings present important information about whether (1) previous claims are generally reliable, and (2) newly suspected relationships are evidenced. However, work on publication bias suggests that journals tend to prefer statistically significant findings, resulting in skewed evidence and inflated meta-analytic effect sizes (Kepes & McDaniel, 2013). Indeed, given that authors are aware of journal preferences, many are unlikely to even submit nonsignificant results for possible publication (i.e., the "file drawer problem:" Rosenthal, 1979).

#### Registered Reports and Results-Blind Review

One way to facilitate the publication of null findings is with "registered" reports (Center for Open Science, 2015a). These studies are preapproved for publication on the basis of their research designs regardless of their findings. This last quality is key, as it minimizes publication risk for scholars and dramatically reduces the value of manipulating data, fabricating results, and unreported HARKing. Further, reviewer advice regarding the proposed study can be provided at the most opportune moment (prior to data collection), avoiding the waste of producing research with illogical theory or fundamental design flaws. Registered Reports have begun to appear in psychology journals such as Social Psychology, Psychological Science, and Perspectives on Psychological Science in partnership with organizations such as the Center for Open Science (Association for Psychological Science, 2015a; Center for Open Science, 2015a). In management, the Journal of Business and Psychology is "trialing" a parallel submission scheme based on the Registered Reports model ("JBP Provisional Acceptance Special Initiative resource site," 2015), and recently published special issues full of null results (Landis, James, Lance, Pierce, & Rogelberg, 2014; Landis & Rogelberg, 2013). Although this initiative enabling the publication of null findings is very welcome, it is an exception that proves the rule.

Alternately, completed studies can be reviewed "results-blind" to reduce bias against papers with null-findings (Newcombe, 1987). Under this scheme, first-round revision decisions are made without the Results or Discussion sections of the manuscript (Kepes & McDaniel, 2013). This procedure helps overcome reviewers' tendencies to base manuscript evaluations on the findings (Greenwald, 1975; Mahoney, 1977), rather than the research questions and design.

Also important, if null results were just as publishable as statistically significant results, there would be much less of an incentive to HARK or manipulate data. As such:

P3c: Adoption of journal policies that enable the frequent publication of null results would reduce (negatively moderate) the otherwise positive relationship between noncredible research practices and career benefits.

If P3c holds, then scholars will engage in NCRPs less. Thus:

P4c: Adoption of journal policies that enable the frequent publication of null results would decrease the prevalence of noncredible research practices.

P5c: Journal policies that enable the frequent publication of null results would decrease the prevalence of noncredible research claims.

#### Open Practice Article Badges

Frameworks for credible science are also being provided by organizations such as the Center for Open Science. For example, journals outside of management have begun using "Open Practice badges." The three badges appear next to the title of an article to indicate whether it has the credibilityboosting features of being (1) Preregistered: Study plan details were registered in an approved, timestamped registration system prior to execution, with results reported in accordance with the plan; (2) Open materials: Indicates the authors have electronically shared the design, measurement details, and study description in sufficient detail to facilitate strict replication; and/or (3) Open data: The codebook (with sufficient description) and original data have been submitted to an approved data repository for subsequent public access and future reanalyses (Center for Open Science, 2015b). The badges are already freely available to journals under a Creative Commons license (with attribution; Center for Open Science, 2015b), and our review found that major publishers of management journals (Sage, Wiley, APA, Elsevier) already have at least one journal (outside of management) currently using these badges.

The badges award positive public recognition to authors who provide credibility-boosting materials for their claims (see also, Association for

Psychological Science, 2015b). In addition, other scholars may feel more confident citing and building on research with credibility-signaling badges. The Open Data and Open Materials badges can also be awarded retroactively, enabling journals to invite and reward authors who provide their data or supplementary materials (e.g., replication protocols) to support the credibility of their prior claims.

However, claims produced through NCRPs are less able to benefit from this status system. As noted above, research by Wicherts, Bakker, and Molenaar finds a link between authors' unwillingness to provide the data associated with their published claims and a "higher prevalence of apparent errors" in reported results, with particularly high degrees of unwillingness when the "errors had a bearing on statistical significance" (2011: 1). As such, Open Practice badges can offer an important signal of the degree of confidence that original study authors have in their claims Thus, the absence of an Open Practice badge can highlight the studies whose claims should be relied upon more tentatively, where replication attempts may be warranted. Given that Open Practice badges provide a mechanism to reward more credible research:

P3d: Adoption of Open Practice badges by journals would reduce (negatively moderate) the otherwise positive relationship between noncredible research practices and career benefits.

If P3d holds, then self-interested scholars will engage in NCRPs less. Thus:

P4d: Adoption of Open Practice badges by journals would decrease the prevalence of noncredible research practices.

P5d: Adoption of Open Practice badges by journals would decrease the prevalence of noncredible research claims.

#### THE GATEKEEPER'S DILEMMA

Although some specific credibility-supportive journal practices highlighted above are recent innovations (crowd-sourcing strict replications, replication protocols, Open Practice credibility badges), calls for journals to reform their policies in the directions described above are not; for example,

publishing null findings and replications, requiring data submissions along with publication acceptance (Ceci & Walker, 1983; Kerr, 1998; King, 1995; Peterson, 1995). Calls for changing the priorities of management scholarship have been made for years to little effect. Nord (1985), for example, recommended devoting 10% of journal space to well-conducted replications some 30 years ago. Around the same time, Mittelstaedt and Zorn (1984:13) noted that replications are difficult to publish due to "the current emphasis on originality, particularly by journal editors and referees," which "probably encourages needless and wasteful product differentiation by researchers who are understandably anxious to be published" (emphasis added). Why haven't more journals adopted credibility-supportive journal practices?

#### The "Citability" Imperative

Here we theorize the as-yet unconsidered cost to individual journals who adopt CSJPs. We propose that journal editors and reviewers are also facing a social dilemma, where what is good for the field (i.e. credibility-supportive journal practices) is bad for an individual journal's prestige. As pointed out by Hollenbeck (2008), at least one factor guiding editors' manuscript decisions is maintaining or improving their journal's impact factor:

"We propose that journal editors and reviewers are also facing a social dilemma, where what is good for the field (i.e., credibility-supportive journal practices) is bad for an individual journal's prestige."

It is critical to remember that ... the editor "has his problems too." He or she is going to be judged by history ... and his or her journal is going to be ranked in terms of its prestige ... as operationalized in terms of citation counts. People who take on such time consuming, difficult, and uncompensated roles typically care deeply about their discipline and are achievement oriented. No one wants to see the impact of the journal he or she is stewarding or his or her discipline sink during his or her reign (p. 17).

Many editorial statements refer to impact factors and, for example, impress upon authors the importance of "being widely cited and contributing to AMJ's impact factor score as a result of earning a high citation count" (Ireland, 2008: 409; see also Rynes, 2007).

Reviewers are also critical gatekeepers in the kinds of scholarship that are published, as editors typically aim to achieve some level of agreement among reviewers about the fate of the manuscript (Hollenbeck, 2008). As reviewers volunteer their labor, many editors are rightfully concerned about overruling reviewers, lest they cease to contribute. As such, the opinions of reviewers also strongly affect the kinds of articles that appear in journals.

Like editors, reviewers are also likely to at least implicitly (if not explicitly) judge manuscripts on the basis of features that contribute to their citability. Three reasons stand out. First, reviewers have been socialized by peers and the review process to value scholarship that has the characteristics that imply high citability. Thus, when asked to recommend which manuscripts should be published, reviewers may fall prey to the naturalistic fallacy (also called the is-ought fallacy): the tendency to infer that the world "ought" to be the way that it "is" descriptively. In our context, the implication is that reviewers will infer that the kind of research they ought to value when reviewing for a management journal is the kind they have been rewarded for producing, and the kind of research that is currently widely prevalent in "high-impact" journals. Second, reviewers' evaluations may reflect explicit instructions from reviewer guidelines or from editors to privilege citability, by emphasizing theory-adding contributions (Bartunek, Rynes, & Ireland, 2006; Pillutla & Thau, 2013) over those that provide strong tests of reproducibility. Third, reviewers may see their role as protecting the journal's status by helping to advise the editor to only publish articles that maintain or enhance the journal's prestige, and therefore, favor the kind of research that is citable.

If citability is important for both editors and reviewers, the question arises: "What kind of research is highly citable?" Two characteristics stand out: theory-adding and statistically supported. First, as noted by Colquitt and Zapata-Phelan (2007), scholarship can add to theory by introducing new concepts or new relationships. Theory-adding scholarship develops claims that are unique compared to prior research (Hambrick, 2007; Mone & McKinley, 1993). In management,

theory-adding research tends to be more highly cited (Beyer, Chanove, & Fox, 1995; Judge, Cable, Colbert, & Rynes, 2007; Kerr, Tolliver, & Petree, 1977). Second, citable claims are statistically supported. There is less reason to cite a study suggesting that a relationship does not appear to exist, and there is widespread bias against publishing null findings (Fanelli, 2011; Greenwald, 1975). In sum, citable research is theory-adding and statistically supported.

In contrast, research is not theory-adding when it confirms what was already thought to be true, or when it "subtracts theory" by indicating that something we thought to be true is not. Such research is unlikely to be cited. For example, if a strict replication identifies a prior finding as potentially erroneous, it may signal the end of a line of inquiry, curtailing future citations. Conversely, if a strict replication confirms the prior study, it seems unlikely to warrant extensive citation, given the norm in management science to cite (only) the first study to make a claim. As such, Hollenbeck advises authors that "straight replication, which is valuable but not urgent in the eyes of most editors... will not be viewed as generating new knowledge nor will it generate a large number of citations—which of course, is a problem for the editor" (2008: 23). Indeed, in marketing, replications and tests of already extant theory are cited less than other types of studies (Hubbard & Armstrong, 1994). As such, it is likely that a journal's citation per paper rate (usually referred to as its "impact factor") will decrease if it publishes retests of prior claims.

Journal gatekeepers' concern for citability renders problematic the range of credibility-supportive journal practices considered previously. Indeed, making the journal policy changes so often advocated, such as publishing more strict replications and null findings, is likely to harm the focal journal's prestige (impact factor relative to other journals) if executed unilaterally.

Similar arguments can also be made for requiring authors to submit their data to repositories upon acceptance for publication. Some authors are understandably reluctant to do so. Data repositories allow other researchers to (a) question the validity of published claims (which could be embarrassing), and (b) to use that data for their own research purposes (perhaps "scooping" an idea of the original authors). Moreover, proper preparation of data such that others can easily engage in reanalysis can be time consuming (Mezias & Regnier,

2007). As such, a journal that makes a policy of requiring that scholars submit their data and a codebook to a data repository upon article acceptance is likely to see a decrease in the number of submissions. Even with multiyear data embargoes, which reduce the chance of being scooped somewhat, the possibility of being shown to have misanalyzed one's data are real (Simonsohn, 2013). Similarly, adopting in-house analysis check policies, publishing refutations, and using Open Science badges can lead to the discovery of errors, which can damage the future citability of claims in the journals that adopt these practices. In sum, journals are likely to receive fewer submissions and fewer citations to the papers they publish if they adopt the credibility-enhancing recommendations that would "change the game."

Further, let us consider what would happen if a lone journal editor was to sacrifice their journal's prestige on the altar of truth. In other words, what happens to the purist journal? Our contention is that the purist journal would only change the incentives for scholars if they are located in an uncompetitive intellectual space; that is, where few other career-enhancing outlets are competing for the same set of submissions and there are high barriers to entry for new journals. In contrast, if a prestigious journal located in a competitive intellectual space opted for "purist" CSJPs, then other nonpurist journals that continue with credibilityreducing practices would simply, over time, assume the mantle of the more prestigious journal. The "game" would not have changed for scholars, only where they submit their research. In such a situation, the purist journal would only serve as a cautionary example to other editors.

In sum, the current emphasis in management scholarship on citability is likely to constrain both journal editors and reviewers (including those who are supportive of credibility-supportive journal practices, but) who hope to maintain or enhance the status of their journals.

P6: Gatekeepers who desire to protect/enhance their journal's prestige will be less likely to adopt credibility-supportive journal practices.

If individual scholars and gatekeepers cannot unilaterally fix the problem of noncredible research, what is to be done? We suggest that collective action is key. Specifically, if other journals in one's area have adopted credibility-supportive journal practices, the focal journal can do so as

well without suffering major decreases in relative status.

"If individual scholars and gatekeepers cannot unilaterally fix the problem of noncredible research, what is to be done? We suggest that collective action is key."

P7: The proportion of "peer" journals that adopt credibility-supportive journal practices reduces (negatively moderates) the negative relationship between gatekeeper's desire to protect/enhance journal prestige and their adoption of credibility-supportive journal practices.

#### SOCIAL DILEMMA SOLUTIONS

Approximately 50 years of research has identified a range of ways to mobilize collective action to solve public-goods social dilemmas. A key distinction in this literature is between structural and motivational solutions (Messick & Brewer, 1983; Zeng & Chen, 2003). Structural solutions change the parameters of a social dilemma by changing the rules of the game to mitigate the costs of "good" behavior for selfinterested actors. Motivational approaches involve changing actors' conceptions of the social environment (rather than the actual characteristics of the dilemma), so actors are more inclined to cooperate in ways that promote the public good (Messick & Brewer, 1983). Both kinds of solutions can work in tandem to address social dilemmas through selfaware collective action (Zeng & Chen, 2003).

#### Structural Solutions

#### Conditional Pledges to Make Adopting Credibility-Supportive Journal Practices Less Costly

In the social dilemma literature, pledges have emerged as a key way to mitigate the costs to actors of addressing social dilemmas (Chen & Komorita, 1994; Chen, 1996). Pledges are simply public commitments by an actor to contribute to the provision of a public good at a certain level. Conditional pledges involve pledges that go into effect only when a substantial proportion of other relevant actors make the same pledge (e.g., once 60% of peer journals have agreed). By making pledges conditional on a critical mass of actors agreeing to change policies toward the public good, actors are only obliged to take steps to promote

collective welfare when they are not substantially competitively disadvantaged by doing so.

An appealing feature of conditional pledges is that, apart from the agreement to take action contingent on proportion of others complying, journal groups have considerable flexibility in the agreement's details, increasing the likelihood of a mutually agreeable outcome. In particular, a conditional pledge is flexible as to (1) which kinds of credibility-enhancing practices leaders agree to adopt (Open Practice badges, registered reports, data-repository requirements); (2) the mode of publishing credibility-boosting research (through special issues, dedicated percent of journal space); (3) the duration of the agreement (e.g., the senior editor's term); and (4) how many journals need to agree for the pledge to be "activated" (e.g., 60% of peer journals).

P8a: Conditional pledges can increase the proportion of peer journals willing to adopt credibility-supportive journal policies.

#### Define Small Peer-Journal Groups

Carefully defining small groups of key actors can be critical for solving social dilemmas for two reasons. First, as noted above, if a journal adopts credibility-supportive journal practices, while other journals do not, then competing journals that do not impose such standards on themselves are likely by comparison to benefit (receive more highly citable submissions, etc.). Thus, avoiding a "race to the bottom" entails identifying the sets of journals that are "competing" with one another to publish the same manuscripts.

A second key reason to define small journal groups is that the transaction costs of communication, coordinating, and negotiating social dilemma solutions can be lower, and thus it can be easier to achieve agreement (e.g., Kerr, 1989). As such, it is important for those seeking to change journal practices (e.g., through conditional pledges) not to define their "journal peers" too broadly (e.g., all management journals). Instead, journal editors are more likely to successfully resolve the social dilemma if they take action as part of a small "social dilemma resolution team."

P8b: Defining small groups of "peer" journals can increase the proportion of journals willing to adopt credibility-supportive journal practices.

In terms of how to decide on the membership of these small journal groups that compete for the same submissions, one option is to identify journals that occupy the same intellectual space. For example, one possibility would be to split journals in the management field into five distinct sets (Felps, van Eck, Waltman, & Meuers, 2014): (1) journals associated with individual-level and small group-level issues; (2) journals associated with human resources and employment relations; (3) journals associated with firm-level issues; (4) journals associated with international business; and (5) general management journals.

Of particular importance is the general management journal set. Because a large number of articles are initially prepared for these high-status journals (e.g., Academy of Management Journal, Organization Science, Administrative Science Quarterly), but ultimately end up at less-prestigious outlets, these journals have a leadership role to play. Their adoption of CSJPs (analysis checking, creating standards for data submission and replication protocols) could have significant positive "follow-on" effects for the specialist management outlets.

#### Reviewers' Boycott/Buycott Pledges

One classic approach for encouraging the adoption of policy solutions from the social movements literature is the boycott, which involves committing to not to deal with an actor until they change their policies as a form of protest (Friedman, 1999; Opp, 2009). Conversely, a "buycott" refers to a commitment to contribute resources to actors who are seen as behaving in a desirable manner (Friedman, 1999). Thus, these structural solutions affect costs and incentives.

Qualified reviewers who are willing to donate their time are a key resource that journals need to operate. Credibility-minded reviewers who feel that the integrity and health of the field requires that journals adopt CSJPs could themselves pledge to (only) review for journals that have signed a conditional pledge by a certain date and/or implement certain credibility-supportive practices. The boycott/buycott serves to (a) create costs for journals that are seen as contributing to problematic incentives for scholars and reduced credibility of management claims, and (b) reward journals willing to adopt credibility-supportive journal practices.

<sup>&</sup>lt;sup>4</sup> Although other factors undoubtedly play a role, we believe that article content is a key driver of which journals scholars submit their manuscripts to (e.g., papers on organizational competitive advantage are usually sent to journals that publish strategy research, not to those that publish Organizational Behavior scholarship).

Organizing a boycott/buycott could be accomplished relatively easily by creating an online form where reviewers pledge to provide their reviewing efforts to journals contingent on their adoption of credibility-supportive journal policies (e.g., agreeing to a conditional pledge) by a certain date, and notifying editors of this initiative. Boycotts and buycotts are classic structural solutions that can be used by a community to increase the number of actors motivated to change.

P8c: The proportion of journals willing to adopt credibility-supportive journal practices would increase if reviewers boycotted/buycotted.

#### Motivational Solutions to Social Dilemmas

Social dilemma scholarship has revealed a number of approaches that can lead prosocially motivated actors to resolve public-goods social dilemmas. Motivational solutions attempt to convince actors to "do the right thing" independent of consequences. Motivational solutions to social dilemmas include (a) encouraging communication among actors in a social dilemma before making decisions about how to behave, and (b) injecting a moral frame.

#### Increasing Multi-Journal Communication

More communication among journal editors before they take steps to resolve a social dilemma is likely to improve cooperation. In a recent meta-analysis, Balliet (2010) reviewed 45 studies looking at the effect of communication on cooperation in social dilemmas. The results revealed a large positive effect of communication on cooperation (d=1.01). Moreover, this effect size was even larger if communication occurred face-to-face (d=1.21), rather than through written messages (d=0.46). Communication about the credibility crisis in management science can increase identification with the group, thereby increasing motivation to contribute to the group's welfare (by implementing credibility-supportive journal practices).

This suggests that although any communication between journals about the credibility crisis is likely to increase adoption of CSJPs, in-person communication is likely to be the most efficacious at inducing cooperation. In particular, we note that although email can be an extremely convenient mode of communication for busy journal leaders, it can also undermine cooperative norms and prosocial behavior as a result of the conflict-exacerbating qualities of email (e.g., minimal social cues, diminished feedback opportunities, ambiguous statements leading

to misunderstandings), which can have cooperationundermining consequences (e.g., deindividuation of others, reduced inhibitions for socially inappropriate behavior, reduced empathy, fewer opportunities to repair disagreements; Friedman & Currall, 2003). Email may be a particularly imperfect medium for discussing issues related to ethical standards, credibility, and professional identity. As such, journal leaders may want to take advantage of opportunities to meet in person as a group (at annual conferences such as the AOM, SIOP, and EGOS conferences) to discuss social dilemmas and opportunities for improvement. Such face-to-face meetings can facilitate the trust needed to improve the effectiveness of subsequent email interactions (Friedman & Currall, 2003).

P9a: Increased multi-journal communication around the topic of research credibility will increase a focal journal editor's desire to produce credible claims.

#### Injecting a Moral Frame

A good deal of research suggests that actors contribute more to a social dilemma when it is framed as a moral situation (Biel & Thøgersen, 2007), especially if they already define themselves as moral people (Aquino, Freeman, Reed II, Lee, & Felps, 2009). In contrast, conceiving of a situation as competitive leads to less cooperation in support of the public good (Pillutla & Chen, 1999). For research gatekeepers, a powerful step toward more credible research is to reframe the situation such that increasing the credibility of management claims is seen as the morally appropriate thing to do. As such, a simple and practical first step for journal reviewers and editors would be to read and share this AMLE Special Section, along with other consciousness-raising calls for reform that describe (a) the credibility issues facing management science, and (b) their consequences for the field, scholars, teachers, students, practitioners, and society (Honig et al., 2014; Kepes et al., 2014; Mezias & Regnier, 2007). Morally framed communications could emphasize the ethical imperative for our field to be credible, such that our claims can be reliably acted upon (cf., Terpstra & Rozell, 1998). For example, in communicating with peers, gatekeepers could emphasize that because most people have a job, management science has the potential to both directly and indirectly improve the productivity and wellbeing of perhaps billions of individuals by ensuring the reliability of our claims.

P9b: If revising journal policies is framed as a moral issue, then a focal journal editor's desire to produce credible claims will increase.

Indeed, social-dilemma scholarship has revealed that many people contribute to social dilemmas even when it is not in their best interest. The rationale for such contributions are a desire to "do the right thing" by contributing to the social good (Messick & Brewer, 1983). As people who volunteer their valuable time to serve as reviewers and journal editors, it is likely that many gatekeepers are already motivated to contribute to the public good. In the case of journal gatekeepers, one key way to "do the right thing" is to produce credible claims by adopting credibility-supportive policies. A surprising amount of change may be possible if gatekeepers were to see enhancing the credibility of management science as a key priority.

P10: Gatekeepers who desire to produce credible claims will adopt credibility-supportive journal practices.

P11: Gatekeepers who desire to produce credible claims will endorse (a) structural and (b) motivational solutions.

Gatekeepers' conception of the appropriate thing to do can be influenced by more than their moral reasoning. Continuing with the theme that gatekeepers are subject to the same tendencies as everybody else, gatekeepers are likely to match the behavior of their peers. This peer influence could be a function of informational conformity (i.e., peers have valuable information about the wise thing to do) or normative conformity (i.e., doing what peers do will allow the gatekeeper to gain acceptance as a respected community member; Deutsch & Gerard, 1955). Due to either or both of these factors, we posit that if a gatekeeper's peers adopt credibility enhancing policies, they are more likely to do so as well.

P12: Gatekeepers' desire to produce credible claims will increase as the proportion of peer journals adopting credibility enhancing policies increases.

#### CONCLUSION

In closing, we have suggested that noncredible research practices (NCRPs) on the part of individual scholars can simultaneously result in noncredible claims in management science and career benefits for individual researchers. We then extend the discussion of credibility issues in management science by arguing that individual scholars and individual journals are both trapped in a social dilemma, where individually rational actions (e.g., publishing claims produced through NCRPs to produce career benefits and journal status) contradict collective welfare (e.g., publishing large amounts of credibility-enhancing science, and adoption of credibility-supportive journal practices).

We suggest that collective action by a substantive proportion of journals nested in small "peer" journal groups can effectively respond to the credibility crisis in management. Drawing on social dilemma research, we offer five kinds of solutions: (1) using conditional pledges to adopt credibility-supportive journal practices while preserving relative journal status; (2) defining small groups of peer journals that compete for manuscripts; (3) reviewer boycotts/buycotts of journals to incentivize journal editors to sign conditional pledges; (4) increasing cross-journal (ideally face-to-face) communication to discuss these issues prior to taking action; and (5) injecting a moral frame into the discussion of journal policies. This framework, along with a bit of institutional entrepreneurship by those committed to credible science, could lead to a paradigm shift in management science (Hubbard, 2015; Hubbard & Lindsay, 2013). We hope that management scholars and leaders in the field will take this opportunity to reimagine the exciting possibilities for management science (and our stakeholders) if we use the powerful levers of aligned incentives, social action, and innovative best practices currently at our disposal.

#### REFERENCES

Aquino, K., Freeman, D., Reed, I. I. A., Lee, K., & Felps, W. 2009. Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, 97(1): 123–141.

Association for Psychological Science 2015a. Registered replication reports. http://www.psychologicalscience.org/index.php/replication.

Association for Psychological Science 2015b. Authors leading the way in open science. http://www.psychologicalscience.org/index.php/publications/journals/psychological\_science/badge-earners.

- Atwater, L. E., Mumford, M. D., Schriesheim, C. A., & Yammarino, F. J. 2014. Retraction of leadership articles: Causes and prevention. *The Leadership Quarterly*, 25(6): 1174–1180.
- Bakker, M., & Wicherts, J. M. 2011. The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43(3): 666–678.
- Balliet, D. 2010. Communication and cooperation in social dilemmas: A meta-analytic review. The Journal of Conflict Resolution, 54(1): 39–57.
- Bartunek, J. M., Rynes, S. L., & Ireland, R. D. 2006. What makes management research interesting, and why does it matter? Academy of Management Journal, 49(1): 9–15.
- Bedeian, A. G., Taylor, S. G., & Miller, A. N. 2010. Management science on the credibility bubble: Cardinal sins and various misdemeanors. Academy of Management Learning & Education, 9: 715–725.
- Benjamini, Y. 2010. Discovering the false discovery rate. Journal of the Royal Statistical Society. Series B, Statistical Methodology, 72(4): 405–416.
- Bettis, R. A. 2012. The search for asterisks: Compromised statistical tests and flawed theories. Strategic Management Journal, 33(1): 108–113.
- Beyer, J. M., Chanove, R. G., & Fox, W. B. 1995. The review process and the fates of manuscripts submitted to AMJ. Academy of Management Journal, 38: 1219–1260.
- Beyer, J. M., Chattopadhyay, P., George, E., Glick, W. H., Ogilvie, D. T., & Pugliese, D. 1997. The selective perception of managers revisited. Academy of Management Journal, 40: 716–737.
- Bhattacharjee, Y. 2013, April 23. The mind of a con man. *The New York Times*, MM44.
- Biel, A., & Thøgersen, J. 2007. Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, 28(1): 93–112.
- Blanton, H., et al. 2009a. Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *The Journal of Applied Psychology*, 94: 567–582.
- Blanton, H., et al. 2009b. Transparency should trump trust: Rejoinder to McConnell and Leibold (2009) and Ziegert and Hanges (2009). *The Journal of Applied Psychology*, 94: 598–603.
- Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. 2015. HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psy*chology, 69(3): 709–750.
- Briner, R. B., & Walshe, N. D. 2013. The causes and consequences of a scientific literature we cannot trust: An evidence-based practice perspective. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 6(3): 269–272.
- Ceci, S. J., & Walker, E. 1983. Private archives and public needs. The American Psychologist, 38(4): 414–423.
- Center for Open Science 2015a. Improving the credibility of scientific research. http://centerforopenscience.org/pr/2014-05-19/.
- Center for Open Science 2015b, January 24. Badges to acknowledge open practices. https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/.

- Chen, X. P. 1996. The group-based binding pledge as a solution to public goods problems. *Organizational Behavior and Human Decision Processes*, 66: 192–202.
- Chen, X., & Komorita, S. S. 1994. The effects of communication and commitment in a public goods social dilemma. Organizational Behavior and Human Decision Processes, 60: 367–386.
- Colquitt, J. A., & Zapata-Phelan, C. P. 2007. Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. Academy of Management Journal, 50(6): 1281–1303.
- Credé, M., & Harms, P. D. 2015. 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior*, 36(6): 845–872.
- Dawes, R. M. 1980. Social dilemmas. *Annual Review of Psychology*, 31: 169–193.
- Deutsch, M., & Gerard, H. B. 1955. A study of normative and informational social influences upon individual judgment. Journal of Abnormal and Social Psychology, 51(3): 629–636.
- Drazin, R., & Kazanjian, R. K. 1990. A reanalysis of Miller and Friesen's life cycle data. *Strategic Management Journal*, 11: 319–325.
- DRYAD 2015. Frequently asked questions: Do I have the option to embargo release of my data? http://datadryad.org/pages/fag.
- Edwards, J. R., Scully, J. A., & Brtek, M. D. 2000. The nature and outcomes of work: A replication and extension of interdisciplinary work-design research. *The Journal of Applied Psychology*, 85: 860–868.
- Fanelli, D. 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, 4(5): e5738.
- Fanelli, D. 2011. Negative results are disappearing from most disciplines and countries. Scientometrics, 90(3): 891–904.
- Felps, W., van Eck, N. J., Waltman, L., & Meuers, J. 2014. The management discipline: A bibliometric and qualitative synthesis. Presented at the Annual meeting of the Academy of Management, Philadelphia, PA.
- Fox, M. 2010. Erratum response to Beamish and Bapuji. *Management and Organization Review*, 6(1): 151–155.
- Francis, G. 2014. The frequency of excess success for articles in Psychological Science. *Psychonomic Bulletin & Review*, 21(5): 1180–1187.
- Friedman, M. 1999. Consumer boycotts: Effecting change through the marketplace and the media. New York, NY: Routledge.
- Friedman, R. A., & Currall, S. C. 2003. Conflict escalation: Dispute exacerbating elements of e-mail communication. *Human Relations*, 56(11): 1325–1347.
- Goldfarb, B. D., & King, A. A. 2016. Scientific apophenia in strategic management research. Strategic Management Journal, 37(1): 167–176.
- Gong, H., et al. 2014. Commentaries and rejoinder on Žeželj and Jokić (2014). Social Psychology, 45(4): 327–334.
- Greenwald, A. G. 1975. Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82: 1–20.

- Hambrick, D. C. 2007. The field of management's devotion to theory: Too much of a good thing? Academy of Management Journal, 50(6): 1346–1352.
- Handelsblatt Business Administration Ranking 2009. Top 100 researchers under 40 ranking. Handelsblatt. http://tool.handelsblatt.com/tabelle/?id=30.
- Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. 2017. Publication bias in strategic management research. *Journal of Management*, 43(2): 400–425.
- Hollenbeck, J. R. 2008. The role of editing in knowledge development: Consensus shifting and consensus creation. In Y. Baruch, A. M. Konrad, H. Aguinus, & W. H. Starbuck (Eds.), Journal editing: Opening the black box: 16–26. San Francisco, CA: Jossey Bass.
- Honig, B., Lampel, J., Siegel, D., & Drnevich, P. 2014. Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? *Journal of Management Studies*, 51(1): 118–142.
- Hubbard, R. 2015. Corrupt research: The case for reconceptualizing empirical management and social science. Newcastle upon Tyne, UK: Sage.
- Hubbard, R., & Armstrong, J. S. 1994. Replications & extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11: 233–248.
- Hubbard, R., & Lindsay, R. M. 2013. From significant difference to significant sameness: Proposing a paradigm shift in business research. *Journal of Business Research*, 66(9): 1377–1388.
- Hubbard, R., & Vetter, D. E. 1996. An empirical comparison of published replication research in accounting, economics, finance, management, and marketing. *Journal of Business Research*, 35(2): 153–164.
- Hubbard, R., Vetter, D. E., & Little, E. L. 1998. Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. Strategic Management Journal, 19(3): 243–254.
- Ioannidis, J. P. A. 2005. Why most published research findings are false. *PLoS Medicine*, 2: 0696–0701.
- Ireland, R. D. 2008. From the editors: Your manuscript's journey through the AMJ review process. Academy of Management Journal, 51(3): 409–412.
- JBP Provisional Acceptance Special Initiative resource site. 2015, January 19. https://jbp.uncc.edu/.
- John, L. K., Loewenstein, G., & Prelec, D. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5): 524– 532.
- Jones, S. R. G. 1992. Was there a Hawthorne Effect? American Journal of Sociology, 98(3): 451–468.
- Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. 2007. What causes a management article to be cited - Article, author, or journal? Academy of Management Journal, 50(3): 491–506.
- Karabag, S. F., & Berggren, C. 2012. Retraction, dishonesty and plagiarism: Analysis of a crucial issue for academic publishing, and the inadequate responses from leading journals in economics and management disciplines. *Journal of Applied Economics and Business Research*, 2(4): 172–183.

- Kepes, S., Bennett, A., & McDaniel, M. 2014. Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. Academy of Management Learning & Education, 13: 446–466.
- Kepes, S., & McDaniel, M. A. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? Industrial and Organizational Psychology: Perspectives on Science and Practice, 6(3): 252–268.
- Kerr, N. L. 1989. Illusions of efficacy: The effects of group size on perceived efficacy in social dilemmas. *Journal of Experimental Social Psychology*, 25(4): 287–313.
- Kerr, N. L. 1998. HARKing: Hypothesizing after the results are known. Personality and Social Psychology Review, 2(3): 196–217.
- Kerr, S., Tolliver, J., & Petree, D. 1977. Manuscript characteristics which influence acceptance for management and social science journals. Academy of Management Journal, 20(1): 132–141.
- King, G. 1995. Replication, replication. PS, Political Science & Politics, 28(3): 444–452.
- Landis, R. S., James, L. R., Lance, C. E., Pierce, C. A., & Rogelberg, S. G. 2014. When is nothing something? Editorial for the null results special issue of Journal of Business and Psychology. *Journal of Business and Psychology*, 29(2): 163–167.
- Landis, R. S., & Rogelberg, S. G. 2013. Our scholarly practices are derailing our progress: The importance of "nothing" in the organizational sciences. *Industrial and Organizational Psy*chology: Perspectives on Science and Practice, 6(3): 299–302.
- Lehrer, J. 2010, December 13. The truth wears off. Is there something wrong with the scientific method? *The New Yorker*.
- Leung, K. 2011. Presenting post hoc hypotheses as α-priori: Ethical and theoretical issues. Management and Organization Review, 7(3): 471–479.
- Mahoney, M. J. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research, 1(2): 161–175.
- Marcus, A., & Oransky, I. 2014. What studies of retractions tell us. Journal of Microbiology & Biology Education, 15(2): 151–154.
- Marcus, E. 2014. Credibility and reproducibility. Cell, 159(5): 965–966.
- Matlack, C. 2013, June 24. Research fraud allegations trail a German B-school wunderkind. *BloombergView*. http://www.bloomberg.com/bw/articles/2013-06-24/research-fraud-allegations-trail-a-german-b-school-wunderkind.
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. 2006. Lessons from the JMCB Archive. *Journal of Money, Credit and Banking*, 38(4): 1093–1107.
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. 2008. Do economics journal archives promote replicable research? The Canadian Journal of Economics. Revue Canadienne d'Economique, 41(4): 1406–1420.
- Messick, D. M., & Brewer, M. B. 1983. Solving social dilemmas: A review. Review of Personality and Social Psychology, 4(1): 11–44.
- Mezias, S. J., & Regnier, M. O. 2007. Walking the walk as well as talking the talk: Replication and the normal science paradigm in strategic management research. *Strategic Organization*, 5(3): 283–296.

- Mittelstaedt, R. A., & Zorn, T. S. 1984. Econometric replication: Lessons from the experimental sciences. *Quarterly Journal of Business and Economics*, 23(1): 9–15.
- Mone, M. A., & McKinley, W. 1993. The uniqueness value and its consequences for organization studies. *Journal of Manage*ment Inquiry, 2: 284–296.
- National Institutes of Health 2015, February 19. *Principles and guidelines for reporting preclinical research*. http://www.nih.gov/about/reporting-preclinical-research.htm.
- Nature. 2015. Recommended repositories: Scientific data. Retrieved from: http://bit.ly/1SCmpjb.
- Newcombe, R. G. 1987. Towards a reduction in publication bias. *British Medical Journal*, 295(12): 656–659.
- Nord, W. R. 1985. Looking at ourselves as we look at others: An exploration of the publication system for organization research. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the organizational sciences*: 76–88. Homewood, IL: Irwin.
- 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(6): 615–631.
- O'Boyle, E. H., Rutherford, M. W., & Banks, G. C. 2014. Publication bias in entrepreneurship research: An examination of dominant relations to performance. *Journal of Business Venturing*, 29(6): 773–784.
- OECD 2007. OECD principles and guidelines for access to research data from public funding. Organisation for Economic Co-Operation and Development. http://www.oecd.org/sti/ sci-tech/38500813.pdf.
- Open Science Collaboration 2015. Estimating the reproducibility of psychological science. **Science**, 349(6251): aac4716.
- Opp, K.-D. 2009. Theories of political protest and social movements. New York: Routledge.
- Oransky, I. 2012. Three papers by German management prof retracted for duplication, statistical issues. *Retraction Watch*. http://retractionwatch.com/2012/07/17/three-papers-by-germanmanagement-prof-retracted-for-duplication-statistical-issues/.
- Orlitzky, M. 2011. Institutional logics in the study of organizations: The social construction of the relationship between corporate social and financial performance. *Business Ethics Quarterly*, 21(3): 409–444.
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., & Kindling, M., et al. 2013. Making research data repositories visible: The re3data.org registry.
- Peterson, M. J. 1995. Community and individual stakes in the collection, analysis, and availability of data. *PS, Political Science & Politics*, 28(3): 462–464.
- Pfeffer, J. 2007. A modest proposal: How we might change the process and product of managerial research. Academy of Management Journal, 50: 1334–1345.
- Pillutla, M. M., & Chen, X.-P. 1999. Social norms and cooperation in social dilemmas: The effects of context and feedback. Organizational Behavior and Human Decision Processes, 78(2): 81–103.
- Pillutla, M. M., & Thau, S. 2013. Organizational sciences' obsession with "that's interesting!": Consequences and an alternative. Organizational Psychology Review, 3(2): 187–194.
- PubPeer 2015. Journal activity. https://pubpeer.com/journals.

- Rasmussen, K. B., & Blank, G. 2007. The data documentation initiative: A preservation standard for research. *Archival Science*, 7(1): 55–71.
- re3data.org 2014, April 14. Registry of research data repositories. http://www.re3data.org/.
- Renkewitz, F., Fuchs, H. M., & Fiedler, S. 2011. Is there evidence of publication biases in JDM research. *Judgment and Decision Making*, 6(8): 870–881.
- Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86: 638–641.
- Rynes, S. L. 2007. Time flies when you're having fun: AMJ, 2005–2007. Academy of Management Journal, 50(6): 1273–1276.
- Scandura, T. A., & Williams, E. A. 2000. Research methodology in management: Current practices, trends, and implications for future research. Academy of Management Journal, 43(6): 1248–1264.
- Schmidt, G. B., & Landers, R. N. 2013. Solving the replication problem in psychology requires much more than a website. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 6(3): 305–309.
- Schmidt, S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2): 90–100.
- Shaffer, J. P. 1995. Multiple hypothesis testing. *Annual Review of Psychology*, 46(1): 561–584.
- 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(11): 1359–1366.
- Simonsohn, U. 2013. Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science*, 24(10): 1875–1888.
- Singh, K., Ang, S. H., & Leong, S. M. 2003. Increasing replication for knowledge accumulation in strategy research. *Journal of Management*, 29(4): 533–549.
- SMJ editors. 2014. Call for papers for a special issue: Replication in Strategic Management. http://bit.ly/23gB4lE.
- Sutton, S. C. 2013. Open access, publisher embargoes, and the voluntary nature of scholarship: An analysis. *College & Research Libraries News*, 74(9): 468–472.
- Terpstra, D. E., & Rozell, E. J. 1998. Human resource executives' perceptions of academic research. *Journal of Business and Psychology*, 13(1): 19–29.
- Tsang, E. W. K., & Kwan, K. 1999. Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24(4): 759–780.
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. 2012. The criterion-related validity of integrity tests: An updated meta-analysis. *The Journal of Applied Psychology*, 97(3): 499–530.
- Wagenmakers, E., Wetzels, R., Borsboom, D., & van der Maas, H. L. 2011. Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3): 426–432.
- Webster, E. J., & Starbuck, W. H. 1988. Theory building in industrial and organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology*: 93–138. London: Wiley.

- Wicherts, J. M., Bakker, M., & Molenaar, D. 2011. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One*, 6(11): e26828.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. 2006. The poor availability of psychological research data for reanalysis. The American Psychologist, 61(7): 726–727.
- Williams, R. 2015. Can't get no reproduction: Leading researchers discuss the problem of irreproducible results. Circulation Research, 117(8): 667–670.
- Zeng, M., & Chen, X. P. 2003. Achieving cooperation in multiparty alliances: A social dilemma approach to partnership management. Academy of Management Review, 28: 587–605.

Eliza Byington (PhD, Rotterdam School of Management) is a lecturer at the University of Sydney Business School. Her research considers a broad array of OB and HRM topics, as well as ways to improve the reliability and accumulation of knowledge in management science, and overcoming the "research–practice gap."

Will Felps (PhD University of Washington) is an associate professor at the UNSW Business School. Reoccurring themes in his research include metascience, ethics, organizational behavior, and measuring knowledge.