

A Call for Openness in Research Reporting: How to Turn Covert Practices Into Helpful Tools

ANDREAS SCHWAB

Iowa State University, Ames, Iowa

WILLIAM H. STARBUCK

University of Oregon, Eugene, and New York University

Research articles often give inaccurate information about how researchers developed hypotheses, analyzed data, and drew conclusions. Published articles sometimes report only some hypotheses that researchers tested, or some statistical analyses that researchers made. Articles often imply that researchers formulated all hypotheses before they examined their data, when in fact they added or deleted hypotheses after they made some data analyses. Indeed, such covert practices are so common that new entrants into management research may think they are correct behavior. Yet, these practices create false impressions about the validity of research and they undermine the openness that ought to create trust among researchers.

Researchers have tried to halt these practices by labeling them “unethical,” but their continued prevalence questions the effectiveness of wholly critical approaches. We propose a constructive path toward reform: advocating honesty about actual research processes by adding discussions of inferences drawn after data analyses. Post hoc data analyses can stimulate important theoretical ideas; running alternative statistical models can deepen understanding of empirical patterns; lack of support for hypotheses can identify incorrect or incomplete theories. The management research culture should encourage these practices. Their negative effects result from the lack of explicit reporting about them.

BIG AND LITTLE LIES IN ACADEMIC RESEARCH

Diederik Stapel rose rapidly to the top of researchers in social psychology. After earning a PhD in 1997, he began to publish frequently in prestigious journals. Not only did his articles deal with topics of current interest to other psychologists and the media, but they often showed that subtle prior stimuli had surprisingly strong effects on later behavior.

We thank Gerard Hodgkinson, Benson Honig, Chengwei Liu, Henrieta Hamilton Skurak, Paul Spector, Daniël Wigboldus, and three anonymous reviewers for helpful comments on drafts of this article. Arthur Bedeian contributed an especially helpful and detailed critique.

In 2009, the Society of Experimental Social Psychology chose him for its “Career Trajectory Award.” He had published 130 articles and 24 book chapters—approximately 10 articles and 2 chapters per year. In 2010, the University of Tilburg appointed him dean of the social and behavioral sciences faculty.

Over the year after his appointment as dean, Stapel’s remarkable achievements collapsed into disaster. After three young researchers voiced suspicions, committees investigated and concluded that at least 76 articles or chapters by Stapel or his students contained data he had faked or manipulated. The University of Tilburg suspended him from

employment. In 2013, the *New York Times* published a long article about Stapel and his research—an article no one wants to read about oneself (Bhattacharjee, 2013).

Stapel's case and some other recent high-profile cases of academic misconduct have received ample attention because they represent intentional and elaborate deviations from ethical norms (Bakker & Wicherts, 2011; Honig & Bedi, 2012; <http://retractionwatch.com/>). The involved individuals were fully aware that they were violating ethical norms. Once discovered, the academic community swiftly condemned and corrected their actions through retractions and ethics investigations. These blatant cases of academic dishonesty are what one could call "big lies."

Big lies represent only a tiny fraction of the misrepresentations in research articles. As Honig, Lampel, Siegel, and Drnevich (2014: 25) observed, "far more common is research conduct that skirts at the edges of what is ethically acceptable." These are "little lies." Statements by researchers, letters from editors to authors, and audits of published studies indicate that little lies are omnipresent in management research. They come in various forms and shapes—and in contrast to discovered big lies—little lies operate below the threshold that triggers strong ethical concerns and sanctions. In their hidden and multifold ways, little lies have had strong corrosive effects for research culture and probably scientific progress. We focus on a few types of little lies in management research that seem to be very common, identifying their detrimental effects, and proposing specific solution strategies.

Some types of little lies, such as not reporting nonsignificant findings or inventing hypotheses after making statistical analyses, have grown so common that many researchers regard them as normal behavior. Editors and reviewers often encourage authors to engage in them during the review process. Like big lies, little lies diminish the trust in research, thereby poisoning academic discourse, public trust, and scientific progress.

The experiences of Candide¹ illustrate some corrosive effects of little lies. He was still a doctoral student when he studied the information content in corporations' annual Letters to Stockholders. He formulated hypotheses based on readings in his sociology minor. Then he collected letters from matched samples of corporations at risk of going

bankrupt and successful corporations that had earlier resembled the failing ones. He was surprised to find no statistically significant differences between the letters from successful and unsuccessful corporations. Other doctoral students and professors proposed additional hypotheses, but these too yielded no statistically significant differences. So, Candide added an interpretation of the findings that pointed out that corporations hire public-relations firms to write the letters, and corporations probably try to minimize or conceal financial problems.

Because several researchers had recently published analyses of letters to shareholders, Candide's faculty advisor urged him to submit this manuscript to a very prestigious journal. Candide was elated when the journal's editor invited him to revise his manuscript. The editor and reviewers said that Candide had not investigated all of the possible differences in letters from the two categories of corporations, so they suggested that Candide should test several more hypotheses. Indeed, they made similar requests three times. Three times the editor asked Candide to revise, and each time the editor and reviewers proposed more hypotheses to test. After the third revision, the editor rejected the manuscript, which now incorporated many hypotheses that the editor and reviewers had proposed but data analyses had not supported. Candide was devastated, but he swallowed his frustration and submitted the manuscript to another highly prestigious journal. Events at this second journal proceeded similarly to the ones at the previous journal. The editor asked for two revisions; each time the editor and reviewers proposed more hypotheses to test; the editor rejected the second revision. By this time, the manuscript incorporated dozens of hypotheses, nearly all of which had originated with editors or reviewers and none of which had yielded a statistically significant coefficient or difference. Candide put the manuscript in a file drawer, which he locked. It has remained there for over 30 years.

Candide's next study examined more than 2,000 letters to shareholders. It received an award for being the best article of the year in a prestigious journal; however, Candide saw these outcomes as evidence that the field of management had poor values. His study had found trivially small differences that attained statistical significance only because of the very large sample, but he needed to publish to gain tenure.

To Candide's disappointment, his third empirical study failed to produce statistical significance . . . at

¹ The individual referred to as Candide has authorized this description of his personal research experiences.

first. This time, he hired a statistician, who tried several additional models and applied several additional statistical techniques to obtain statistically significant results. A prestigious journal published this article.

The foregoing experiences left Candide feeling deceitful and disillusioned. He knew that his published articles did not accurately reflect his research processes. The articles did not reveal that he had added hypotheses during the review process or after making statistical analyses; they only reported statistically significant models, and they did not explicitly discuss the dependence of the results on large samples and exploration of alternative statistical models. He felt his manipulations had produced findings that were not trustworthy. He had also observed that other management professors appeared to have adopted methodologies in this opportunistic way to achieve publication, rather than to discover or validate knowledge. He vowed to do no more quantitative research.

COVERT RESEARCH PRACTICES

We first identify some of the prevalent but covert research practices that Candide encountered. Methods scholars have long identified these practices as deceptive and have labeled them unethical. Standard statistics textbooks instruct readers to avoid them (Mazzola & Deuling, 2013). Still, scholarly insiders keep pointing out their apparent prevalence. In an anonymous article in the *Journal of Management Inquiry*, for example, an established researcher revealed in detail on how he or she engaged in dishonest reporting of how a research team had arrived at their results. The author in hindsight described the outcome as "What we wrote in the article was a lie. It amounted to academic dissembling even though I knew it was commonly done" (Anonymous, 2015: 214).

Banks, Rogelberg, Woznyj, Landis, and Rupp (2016b) examined 64 business studies and inferred that 91% of these studies showed evidence of covert, undesirable practices in the conduct or reporting of research.

Because covert research practices warp and constrain scientific progress, researchers need to discuss how to correct them. Past discussions have labeled the practices discussed in this article as unethical and tried to enforce norms that discourage researchers from engaging in them. The continuing prevalence of these practices, however, questions the effectiveness of a wholly critical approach. A

positive approach may be more effective. We argue that the corrosive effects of these covert practices result primarily from concealing their use and that similar research practices can create pathways to deeper understanding. Hence, we propose ways the research community can support comprehensive and complete reporting, thereby reinforcing a fundamental ethical norm: honest and accurate reporting (Merton, 1973). Even more important, we urge researchers to improve and develop these practices to fully exploit their potential to support useful causal inferences. Researchers can turn these currently corrosive research practices into helpful tools.

Three Important Types of Little Lies

Selective Reporting of Hypothesis Tests

Management researchers claim empirical support for more than 90% of the hypotheses they test (Bergh, Sharp, Aguinis, & Li, 2015). This incredible success rate is much higher than would be expected considering the reported measures of significance and sample sizes (Kepes & McDaniel, 2013; Lovell, 1983; Simmons, Nelson, & Simonsohn, 2011). This success rate is even more astonishing because most researchers claim they tested innovative new theories or substantial extensions of prior theories, and they never portray their studies as replications (Pfeffer, 1993; Siler & Strang, 2016). Related fields of social science have claimed similarly implausible success rates. After analyzing multiple psychology articles in *Science*, Francis, Tanzman, and Matthews (2014) estimated that 83% of these articles had claimed success rates that were very unlikely. An effort to replicate 100 psychology studies indicated that although 97% of the original studies reported statistically significant effects, only 36% of the replicated studies did so, and the effects observed in replicated studies were about half as large as those originally reported (Open Science Collaboration, 2015; Simonsohn, 2016). Based on 300 articles in prominent strategic management journals, Goldfarb and King (2015) estimated conservatively that about 25–40% of the published claims of statistical significance are actually false. Such audits strongly suggest that researchers or editors do not publish studies that report null findings (Kepes, Banks, McDaniel, & Whetzel, 2012). After surveying 52 authors of articles in a prominent journal, Siler and Strang (2016) stated that papers that challenge a theoretical perspective face distinctly higher levels of criticism and change

requests during editorial review. Research suggests that null results disappear not only because reviewers and editors reject studies, but also because researchers do not submit such articles, and they drop hypotheses that do not receive statistically significant support (Bedeian, 2003). Consequently, scholars commonly assume that published articles do not describe all of the conducted hypothesis tests, and they doubt the accuracy of reported statistical indicators for hypothesis tests.

This nonreporting of null results seems to be partly a consequence of misinterpretation of statistical significance. A finding that is not statistically significant may be practically important, even very important. Yet, many researchers act and speak as if they can ignore all findings that are not statistically significant (Hubbard & Lindsay, 2013; McShane & Gal, 2016). This behavior creates false impressions about the generality and validity of theories by understating the importance of situational factors and sample sizes. Indeed, as the story about Candide illustrates, differences or effects may be socially or theoretically important precisely because they are very small. The public and legal institutions regard pharmaceutical companies as acting unethically when they suppress tests that show drugs to have weak or no effects, and statistical significance is generally an unreliable indicator of the importance of phenomena because it takes no account of costs or benefits for different stakeholders (Schwab, Abrahamson, Fidler, & Starbuck, 2011; Hubbard, 2015).

Meta-analysis has created new opportunities to aggregate findings from multiple studies and to investigate the consistency of effects across studies (Cumming, 2011; Schmidt & Hunter, 2014). However, accurate meta-analyses require complete records of all studies. Consequently, a bias against publishing nonsignificant or small effects creates severe problems for meta-analyses (Kepes et al., 2012; Biemann, 2013). Researchers have to search for unpublished tests, and they are unlikely to find results for all conducted studies. Hence, the current editorial bias creates severe challenges for meta-analyses.

A research culture that refuses to disconfirm bad hypotheses fills journals and textbooks with "truths" that actually lack scientific support. Journals that publish only articles that confirm hypotheses create an enormous pressure on researchers to find confirmation—especially in a "publish or perish" environment. LeBel, Campbell, and Loving (in press) highlight how the current incentive structure in academic research impedes open reporting, data sharing, and replication. Hence, researchers

engage in various practices to increase their odds of supporting hypotheses. Two of these practices, which appear to be very prevalent, are HARKing and *p*-Hacking.

HARKing: Hypothesizing After Results Are Known

Empirical articles in management journals typically start with elaborate descriptions how the researchers derived hypotheses from existing theories and prior empirical studies. Next, articles claim to report rigorous empirical tests of these formal hypotheses: tests that involve correlation, regression analyses, and statistical significance tests. This structure implies a purely deductive chain of reasoning in which the researchers supposedly derived all current hypotheses from findings in prior studies. However, evidence strongly suggests that this is not how the researchers actually conducted the studies, and the differences between what researchers say they did and what they actually did are not minor. For example, anonymous surveys of authors and editors indicate that authors often select and formulate the hypotheses after or during data analysis (Bedeian, Taylor, & Miller, 2010).

When researchers investigate patterns in their data and then start formulating hypotheses that explain these patterns, they are HARKing (Kerr, 1998). Researchers are also HARKing when they start with broad, general hypotheses, then drop not-supported hypotheses, and develop rationales for hypotheses they inferred from data analyses. And, researchers are HARKing when they amend their original hypotheses in response to data analyses. They may, for example, replace a monotonic hypothesis with a curvilinear hypothesis or replace a two-tailed test with a one-tailed test.

An especially troublesome form of HARKing occurs when journal editors or reviewers advise authors to add to or modify their original hypotheses. Obviously, editors and reviewers know the outcomes of researchers' analyses before they propose alternative explanations, theories and tests, which makes statistical significance tests of their proposals invalid. Even worse in a "publish or perish" environment, researchers are likely to see the "suggestions" of editors and reviewers as demands they must satisfy, and the (hindsight) rationalizations proposed by editors and reviewers as inferences they must draw (Bedeian, 2003). There is no way for editors or reviewers to intervene without invalidating the premises of deductive theorizing.

Furthermore, when editors or reviewers propose that authors add or modify their (supposedly deductive) hypotheses, they create an impression that such behavior is ethically correct.

HARKing makes management theories appear more effective than they are. The propositions of management theory are not only plentiful, they usually occur in mutually contradictory sets, because it is impossible to spell out all of the conditions under which each proposition is valid, so the limitations of each proposition evoke other propositions that describe the consequences of alternative conditions (Schwab & Starbuck, 2016). One result of this plethora of theoretical propositions is that researchers have to choose among contending alternative theories, which is much easier after the researchers obtain findings in a specific situation. Hindsight creates the illusion of powerful theories.

All forms of HARKing also increase the probability of obtaining statistically significant results, and hence of achieving publication (Bosco, Aguinis, Field, Pierce, & Dalton, 2016). They also increase the probability of basing generalizations on idiosyncrasies of specific samples. Thus, HARKing helps explain the strangely high rate with which studies support proposed hypotheses and the low rate at which later studies reproduce earlier findings. In a survey of faculty from PhD-granting management departments, 92% of the respondents reported that they knew faculty who developed hypotheses after they saw their results (Bedeian et al., 2010: 716).

p-Hacking and Best-Model Reporting

p-Hacking involves running multiple statistical tests, but reporting only some of those tests. Modern statistical software facilitates such experimentation: Researchers can change models easily and obtain results in seconds.

Most published measures of statistical significance grossly misrepresent odds of finding statistical significance (Bedeian, Sturman, & Streiner, 2009; Peach & Webb, 1983). Conventional measures of statistical significance assume that only one estimation occurs. Researchers can compute *p* values that allow for multiple estimations if they specify all intended estimations before making any of them (Lovell, 1983), but management researchers do not report doing this. If researchers continue to make estimations in an exploratory way until they get results they like, or if journal editors or reviewers advise authors to make additional estimations, statistical significance tests and *p* values are even less

meaningful than they usually are—probably much less meaningful—because their formulas assume only one estimation. Indeed, simulations show how easy it is to “discover” statistically significant relationships by searching at random through the kinds of data that management researchers analyze (McWilliams, Siegel, & Teoh, 1999; Simmons et al., 2011; Webster & Starbuck, 1988). An article in *Science* offers online access to simulations that allow researchers to experiment and to develop better intuition about the threats of *p-Hacking* (Aschwanden, 2015).

Open-ended exploratory estimations invite misleading inferences about theories’ usefulness. If researchers do not report models that did not support their initial hypotheses, their articles create false impressions about the validity of those hypotheses (Biemann, 2013). In the survey of faculty from PhD-granting management departments cited above, 78% of the respondents said they knew professors who had “selected only those data that support a hypothesis and withheld the rest” (Bedeian et al., 2010: 716).

Because publication affects their job security, researchers are highly motivated to avoid abandoning studies (Miller, Taylor, & Bedeian, 2011). If initial tests of their hypotheses do not yield statistically significant results, many researchers explore alternative models and data configurations to find statistical significance. Widespread *p-Hacking* is a very likely explanation for the excessive success rate of hypothesis tests in published research.

In summary, incomplete reporting of hypothesis tests, HARKing, and *p-Hacking* are likely explanations for the implausible success rate of published hypotheses. These practices both distort evidence about the usefulness of theories and undermine confidence in the conclusions reported (as summarized in Figure 1). The undesirable properties of these practices are well-established in the social sciences and the corresponding methods literature (Banks et al., 2016a; Kepes, Bennett, & McDaniel, 2014; Landis & Rogelberg, 2013; Schwab et al., 2011; Simmons et al. 2011; Starbuck, 2016a). Decades of prior publications have discussed different aspects and implications of these practices (e.g., Greenwald, 1975; Rosenthal, 1979; Sterling, 1959). Professional associations and journal publication guidelines have broadly classified them as unethical (e.g., AOM Ethics Education Committee, 2011; American Psychological Association, 2010). Still, these covert practices have persisted.

The distortions and errors caused by incomplete reporting of hypothesis tests, HARKing and *p-Hacking* are probably large, and management theories are

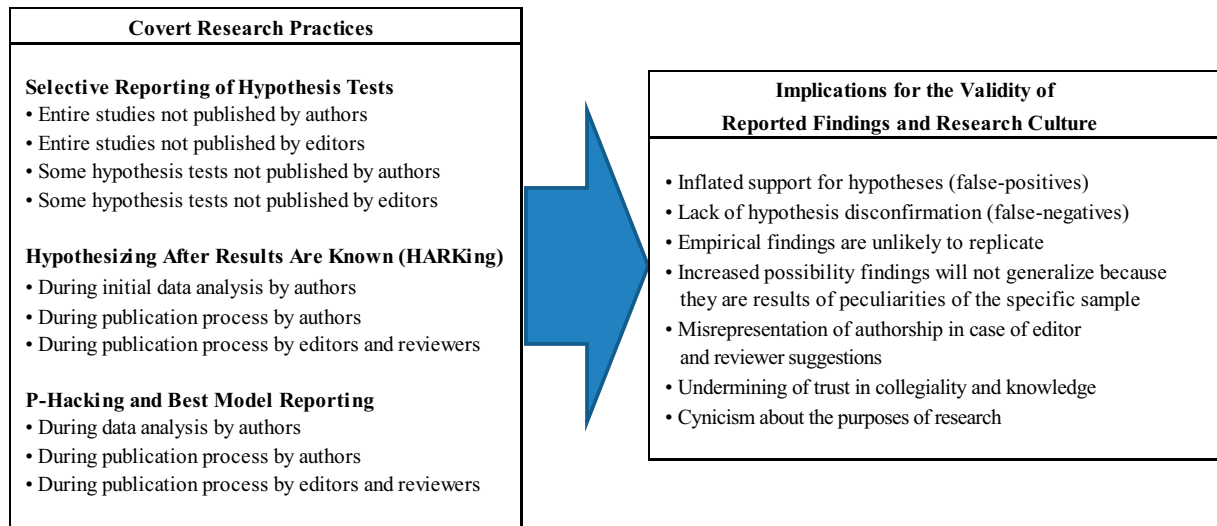


FIGURE 1
Covert Research Practices

probably much less useful than published articles claim. In a recent survey, 50% of management researchers admitted that they selectively reported hypotheses based on statistical significance and portrayed post hoc hypotheses as deductive empirical tests (Banks et al., 2016a). For psychology researchers, John, Loewenstein, and Prelec (2012) reported that researchers self-reported the following prevalence rates: failure to report all dependent variables 78%, collecting more data after seeing if results were significant 72%, stopping data collection after achieving the desired result 36%, selectively reporting studies that worked 67%, excluding data after looking on impact of doing so 62%, and claiming to have predicted unexpected findings 54% (Fiedler & Schwarz, 2016). O'Boyle, Banks, and Gonzalez-Mulé (2014) made another suggestive calculation. In a study of doctoral dissertations in management and psychology, they found alterations as the dissertation research moved toward publication. The alterations included dropping of statistically non-significant hypotheses, adding statistically significant hypotheses, reversing the directions of hypotheses, deleting or adding data after hypothesis tests, and deleting or adding variables. As a result, the ratio of supported to unsupported hypotheses more than doubled.

THE POTENTIAL VALUE OF RESEARCH PRACTICES THAT ARE NOW COVERT

Critics of these covert practices have mainly been urging researchers to try to prevent HARKing or p-Hacking and urging journals to publish high-quality studies whether or not they obtain statistically

significant results (AOM Ethics Education Committee, 2011; Kerr, 1998). This approach has had disappointingly small effect, and it seems likely to continue to fail. Pressures to publish and deeply ingrained practices pose enormous challenges (Orlitzky, 2012). There is no conclusive way to verify whether researchers engaged in these practices, and the proposed remedies offer no incentives to motivate changes in behavior.

Therefore, we propose a radically different approach. HARKing and p-Hacking should become useful investigative techniques. These analyses currently cause harm mainly because significant fractions of articles describe research processes deceptively or incompletely (Fanelli, 2013; Sijtsma, Veldkamp, & Wicherts, 2015; Wigboldus & Dotsch, 2016). These articles misrepresent inferences drawn from data analyses as having been deduced a priori from previous studies or theories. These deceptions in combination with publishing only statistically significant results overstate the correctness and specificity of preexisting theories and understate the new learning made possible by data analysis. Explicit, precise, comprehensive, and honest reporting about research practices is crucial for interpreting findings and for creating a culture of mutual trust (Bem, 2004). And, it is more useful to make dishonesty unnecessary than to try to detect and punish it (Sijtsma, Veldkamp, & Wicherts, 2015).

The Potential Value of Results That Are Not Statistically Significant

There are two problems with a journal policy that rejects manuscripts because they do not report having

found statistical significance. First, when high-quality studies cannot support key hypotheses of well-known theories, these failures should be important information. Researchers need to know that theories have weaknesses, possibly due to requirements that deserve further investigation. Failure to report small relationships distorts cross-study comparisons, such as meta-analyses. There is much evidence that editorial evaluations are unreliable, so reviewers and editors should focus on trying to improve the clarity of research articles and on posing questions that they had as readers (Starbuck, 2016a, 2016b). Second, statistical significance is an unreliable criterion for judging the importance of observed effects. After watching many decades of troubling behavior by researchers, the American Statistical Association Board of Directors (2016) has published a warning against the use of statistical significance or *p* values to justify binary decisions about what is important and what is not. The basic issue is that these indicators are sensitive to the peculiarities of specific samples; repeated samples from the same population may yield very different significance indicators (Cumming, 2011). Readers of research articles will be better able to evaluate findings if the articles conceal nothing and state confidence intervals for all parameter estimates instead of statistical significance.

The Potential Value of HARKing

To develop additional hypotheses based on the available data is not inherently bad. When detectives arrive at the scene of a crime, they try to develop hypotheses about what events occurred, and possibly why these events occurred. Very large fractions of all scientific research and knowledge have begun as conjectures derived by observing data. Indeed, it is irrational that management scientists place extreme emphasis on hypothesizing on the basis of previous studies to the neglect of hypothesizing on the basis of data (Locke, 2007).

Deductions from existing theories and prior studies are only a small part of research; discovering empirical patterns through data analysis is equally important. Unexpected and accidental discoveries have frequently propelled science into new ways of thinking. The fact that existing theories and prior studies did not lead researchers to these discoveries does not make the discoveries irrelevant or unimportant. To the contrary, it is interesting and important when new data generate ideas for new hypotheses, studies, and theories.

When researchers claim dishonestly that they predicted their discoveries deductively, they create

two problems. First, the conventional statistical metrics assume that the data comprise a random sample. When researchers use a sample as the basis for deriving hypotheses, they must no longer regard those data as a random sample when evaluating the derived hypotheses. The derived hypotheses remove the accidental, haphazard quality from that sample. Thus, the researchers cannot base "tests" of the derived hypotheses on an assumption that the sample is still random, and inferences about the studied population, such as *p* values or confidence intervals, are no longer valid. Because all statistical tests assume random sampling, no statistical tests exist for hypotheses that derive in part from properties of the collected data. To test such derived hypotheses, researchers need to obtain new random data. Second, by attributing related inferences to theories, researchers overstate the usefulness and generality of those theories. The diversity of management studies and samples creates a complex mixture of partially conflicting, partially distinct conjectures about the studied phenomena. Rarely or never do researchers encounter situations where a single dominant theory offers clear and strong predictions that apply without only-if requirements. Instead, management researchers usually face a mixture of alternative theories and prior findings with a variety of ill-understood boundary conditions that might apply more-or-less to their own studies. This creates substantial challenges for researchers to identify the most appropriate and promising theory-based hypotheses. These multiple theories, diverse prior findings, and potential but vague boundary conditions create a temptingly heterogeneous pool from which to pick hypotheses. Retrofitting hypotheses to data creates an appearance of support for these hypotheses that vastly overstates their actual ability to make predictions about new samples of data from the same population.

If researchers discover new hypotheses after or during data analysis, they should report such observations as inferences, conjectures, or discoveries. Likewise, researchers should identify hypotheses and proposed models that originated with editors or reviewers as having come from those people after submission to the journal. Honest reporting enables readers to recognize and account for the exploratory nature of these observations. Honest reporting also fosters a broader awareness of the contributions made by inductive and abductive reasoning.

Three types of reasoning—deductive, inductive, and abductive reasoning—can all serve useful roles

in the production of science. The current culture in management research greatly underestimates the value of inductive reasoning, in which researchers derive propositions from their analyses of data. Inductive researchers gather data, analyze the data for patterns, and formulate hypotheses or theories about the observed patterns. The discovery of these patterns may be as important as hypotheses and theories. Analyses of massive databases often rely on inductive reasoning: Some companies, for example, have been able to reduce their inventories substantially by discovering and allowing for different purchase patterns in different locations. Abductive researchers start with the assumption that they have seen only portions of the data that might exist, so the entire universe of data might include phenomena that no one has yet observed. Thus, abduction involves imagination, creativity, and logical extrapolation. Albert Einstein's theories about the structure of the universe exemplify abduction, as did Herbert Simon's conjectures about the future use of computers to simulate human thought. The history of scientific progress includes many important discoveries originating from induction or abduction. The most promising solution to the negative effects of HARKing is to encourage researchers to study and apply inductive and abductive reasoning publicly and with pride.

“The most promising solution to the negative effects of HARKing is to encourage researchers to study and apply inductive and abductive reasoning publicly and with pride.”

The Potential Value of p-Hacking

Running multiple alternative models to probe for patterns in data, including the robustness of these patterns, is generally useful (Wigboldus & Dotsch, 2016). It makes sense to exploit the ease with which modern statistical software packages can examine alternative theories or alternative versions of a fundamental model. Researchers who engage in p-Hacking run multiple models but they omit reporting some models, and they report other models as if they had hypothesized the reported effects in advance rather than as the (possibly surprising) discoveries of exploratory data analysis.

Data are a “black box,” a term that denotes a system having unobservable inner workings. The analytic challenge is to draw inferences about what happens inside a black box. By manipulating inputs systematically and observing the corresponding changes in outputs, researchers can learn about the inner workings of a black box. In data analysis, researchers run alternative models and use alternative statistical procedures to develop a deeper understanding of the relationships among variables in the data set. This resulting deeper understanding adds credibility to inferences about systems that the data describe. Prior hypotheses always leave some variance unexplained, so there is always more that researchers could learn about the data or the studied situation. Running additional models can provide information on what features of the model or empirical setting change the variance explained and by how much. Hence, researchers should always run multiple model configurations and statistical procedures to discover their implications. Would modified hypotheses be more effective? Do other moderating or contingency variables warrant consideration? Do the data have peculiarities that raise questions about the usefulness of generalizations? Hence, practices currently used to support p-Hacking can transform into valuable research tools (Wigboldus & Dotsch, 2016). Instead of discussing how to prevent p-Hacking, researchers should discuss how to systematically perform exploratory, incremental, and iterative multimodel analyses and how to communicate the related findings. As researchers increasingly exploit massive data sets, both the need and the opportunities for exploratory investigations increase. Large data sets may also facilitate testing the predictions of models discovered through exploratory data analysis.

“Instead of discussing how to prevent p-Hacking, researchers should discuss how to systematically perform exploratory, incremental, and iterative multimodel analyses and how to communicate the related findings.”

TURNING LITTLE LIES INTO FORTHRIGHT AND USEFUL PRACTICES

Little lies are only “small” in the sense that they have quiet tolerance. The threat they pose to

management research is large. Studies of research articles in prominent journals collectively indicate that roughly half of the claimed findings are actually false or unreproducible (Open Science Collaboration, 2015; Hubbard, 2015; Bosco et al., 2016; Goldfarb & King, 2015). With such poor credibility, it is difficult to see how management research can make useful contributions.

Raising the credibility of management research is both a very difficult challenge and a very important goal that many stakeholders should support. The challenge is very difficult because so many researchers engage in HARKing and p-Hacking while also keeping them covert. These practices have become deeply engrained in actual research activities in spite of efforts to eliminate them by labeling them as unethical. However, such efforts to constrain use of these practices through public declarations have not only been ineffective, they have supported the notion that their use should be covert.

We have observed the following behaviors firsthand. Very few researchers discuss HARKing and p-Hacking openly; note that the author of one explicit description of these practices asked to be called "Anonymous" (2015). Research teams discuss HARKing only among themselves or with trusted friends. When speaking privately with trusted elders, doctoral students sometimes voice their discomfort with and confusion about HARKing and p-Hacking, but the students do this cautiously and in quiet voices. Although courses and readings tell the students that these practices are dishonest, the students say they observe their professors using them, and say professors have advised them to engage in such practices. Seminar audiences interpret public questions about HARKing and p-Hacking as accusations of unethical behavior—even when questioners speak very diplomatically and sympathetically. Such questions induce presenters to offer reassurances that they obviously would never do such things. Hence, such questions create rather awkward moments. An inability to discuss the issues publicly also creates substantial challenges for collective methodological change.

Professional associations, publishers, universities, journal editors, methodology teachers, and individual researchers all benefit from the appearance that management research is scientific, so all of these entities should have strong motivation to increase the credibility of management research. For a respected professional association or university to support change would be very helpful.

Unfortunately, over the last half century, these entities have repeatedly demonstrated allegiance to current methodological practices and resistance to efforts to reform them. Professional associations and universities have ignored or attempted to play down reform proposals that might upset many scholars, especially more prominent scholars. The prevalence of little lies testifies that they are essentially not the actions of individual researchers but the actions of a social system that tells researchers what to do, and professional associations and universities are the organized public faces of this social system.

A study of the early history of computer simulation may have said something profound about how research practices change. Starbuck and Dutton (1971) classified simulation studies according to how much effort they devoted to validating their assumptions, using realistic input data, and comparing outputs with data about actual events. Studies that had little empirical validation declined gradually over time, and studies that had far more empirical validation increased gradually over time. However, this evolution did not occur because individual researchers changed their methodological practices. About 40% of researchers continued to apply the same methodological practices as in their prior studies, and about 55% of researchers devoted even less effort to empirical validation in their subsequent studies. Standards for empirical validation rose because new adopters of this methodology set higher goals than their predecessors.

Effective initial change efforts are much more likely to come from journals, methodology teachers, and researchers themselves. Quite a few journal editors have attempted to reform research practices, even in the face of strident protests from authors; many methodology teachers teach what they believe to be right instead of what has been traditional; and many individual researchers have, like *Candide*, used and advocated research practices that deviated from widespread patterns.

How Journal Editors Can Help

In 2016, the editors of the *Strategic Management Journal* declared that their journal welcomes replication studies and studies with nonresults: It will no longer publish "papers that report or refer to cutoff levels of statistical significance (p-values)," and authors should "explicitly discuss and interpret effect sizes" (Bettis et al., 2016: 260).

Journal editors can also stimulate profound behavioral changes related to HARKing and p-Hacking. They can ask for explicit comparisons of alternative models to discourage the reporting of only a single best model, they can ask for probability corrections when researchers test multiple alternative models that they proposed before making analyses, and they can require every article to include a section that discusses discoveries that the researchers did not predict before they gathered data for this study. Particularly important, when editors or reviewers suggest that authors add or modify their hypotheses, the editors should state clearly that (a) the articles should attribute these changes to the editors or reviewers, and (b) the articles should describe these changes as having occurred after the data were analyzed.

Management is not the only field trying to confront and deal with covert research practices. Table 1 outlines some current experiments with methodological changes in management, medical science, and psychology. Such experiments sometimes establish new behavioral patterns, but they also sometimes fail. Fidler (2005) found that authors who obeyed journals' requirements to report effect sizes nevertheless discussed their findings in terms of statistical significance, and Chang and Li (2015) inferred that requirements by economics journals to make data and code public had been ineffective because they lacked active enforcement.

Arguing that editors may see risk in adopting new practices to make management research more credible, Byington and Felps (2017; this issue) recommend that editors form coalitions to jointly change editorial policies. Such collective initiatives promise to increase the perceived legitimacy of changes and lessen risks related to deviance, and enlisting editors in a coalition seems substantially easier than convincing an entire professional association to change. For example, 10 journals have jointly offered to implement preregistration of empirical studies (*Journal of Business and Psychology*, 2016).

How Methodology Teachers Can Help

Wasserstein and Lazar (2016) reported that George Cobb had challenged a forum in the American Statistical Association with two question-answer pairs:

Question: Why do so many colleges and grad schools teach $p = .05$?

Answer: Because that's still what the scientific community and journal editors use.

Question: Why do so many people still use $p = 0.05$?

Answer: Because that's what they were taught in college or grad school.

As Cobb indicated, many methodology teachers teach what they believe management journals demand and what management researchers expect rather than what they believe to be useful and methodologically correct (McShane & Gal, 2016). It does make sense to prepare management doctoral students for the wide prevalence of null-hypothesis significance tests (NHSTs) and consequent binary thinking. However, this preparation should include explanations of frequent misunderstandings and misinterpretations of NHSTs, and methodology courses should offer students alternative ways to analyze and interpret data.

"...when editors or reviewers suggest that authors add or modify their hypotheses, ... (a) the articles should attribute these changes to the editors or reviewers, and (b) the articles should describe these changes as having occurred after the data were analyzed."

Table 2 lists several topics that methodology courses should discuss, but often spend too little time on. All these relate broadly to the costs and benefits of making models and theories more simple or more complicated. Humans find it difficult to reason with models that involve more than two or three variables, and they tend to convert continuous gradations to dichotomies, but analytic models that drop less important variables may exaggerate the importance of the retained variables and misrepresent the complexity of studied situations. Whether it is useful to incorporate many variables in models depends on whether analysts want to develop detailed understanding of specific samples, including idiosyncrasies that are unlikely to occur in new data. For generalization or prediction beyond specific data, simpler models are usually more accurate. Obviously, the testing of deductive hypotheses is only one use of statistics, and possibly not the most important use: Methods courses also should discuss analytic approaches for induction and abduction. Published research articles very frequently apply statistical formulas that require randomly selected data to samples that are not random in one way or another. For example, researchers might use conventional statistical formulas to describe data gathered from all

TABLE 1
Methodological Change Efforts in Management, Psychology, and Medical Research

Focus	Management	Psychology	Medical research
Transparency	<ul style="list-style-type: none"> Academy of Management provides a Code of Ethics (2006) that requires members to report comprehensively all findings. Management Science has a data submission requirement. All other management journals on the FT 45 list have no stated data access or replication policies (Jensen, 2015). Meta-analysis is increasingly embraced with an explicit focus on including disconfirming results and unpublished data. <i>Strategic Management Journal</i> (Bettis et al., 2016) encourages replication studies and reporting nonresults. Academy of Management created a new journal, the <i>Academy of Management Discoveries</i>, which focuses on the publication of inductive, abductive, and replication studies.¹ 	<ul style="list-style-type: none"> American Psychological Association's publication guidelines (2010) recommend comprehensive reporting and data sharing. A few journals have mandatory data sharing and some experiment with data archiving (e.g., <i>Journal of Applied Psychology</i>, <i>Journal of Consumer Psychology</i>; Jensen, 2015). Several journals encourage meta-analysis with a focus on including results from unpublished research. Association for Psychological Science and <i>Perspectives on Psychological Science</i> journal are encouraging replication studies. 	<ul style="list-style-type: none"> Cochrane Foundation coordinates the systematic and centralized sharing and analysis of comprehensive research data (www.cochrane.org). Strong societal interests and public scrutiny enforce high levels of transparency (e.g., law suits; FDA investigations). Replication, meta-analysis, and accumulation of research evidence across studies are strongly embraced with an explicit focus on including disconfirming results and unpublished data.
Ethical norms and criticism	<ul style="list-style-type: none"> Academy of Management provides video guidelines that criticize selective reporting, HARKing, and p-Hacking; but propose no enforcement mechanisms. <i>Strategic Management Journal</i> (Bettis et al., 2016) requires discussion of effect size and no longer accepts papers that only report p-value cut-off levels. 	<ul style="list-style-type: none"> American Psychological Association's publication guidelines (2010) criticize selective reporting, HARKing and p-Hacking, but propose no enforcement mechanisms. American Psychological Association's publication guidelines (2010) recommend focus on effect size, confidence intervals, and meta-analysis instead of single statistical significance tests. 	<ul style="list-style-type: none"> Strong societal interests and scrutiny combined with potential for severe negative repercussions (e.g., law suits², loss of professional certification and reputation). Centers at highly respected universities and nonprofit organizations are advocating and promoting change in research methods, including a focus on effect size, confidence intervals, and meta-analysis (e.g., Meta-Research Innovation Center at Stanford (METRICS); Society for Clinical Trials, John Hopkins Center for Clinical Trials and Evidence).
Prevention			

(table continues)

TABLE 1
Continued

Focus	Management	Psychology	Medical research
"Prevention (cont.)"	<ul style="list-style-type: none"> • A few journals are experimenting with registered-study review processes in which researchers submit research proposals to journals for documentation and potentially review before they actually collect data.³ 	<ul style="list-style-type: none"> • A few journals are experimenting with registered-study review processes in which researchers submit research proposals to journals for documentation and potentially review before they actually collect data.³ 	<ul style="list-style-type: none"> • ClinicalTrials.gov represents a large and successful initiative to register and archive clinical trials to minimize publication bias and encourage aggregation of findings across studies (online searchable database, worldwide scope).

¹ In 2017, *Human Resources Management Review* will publish a special issue about effective ways to draw inductive and abductive inferences from data.

² In 2004, for example, the State Attorney General of New York filed a suit against GlaxoSmithKline, the maker of Paxil, for concealing clinical trial studies that indicated their drug Paxil was ineffective for pediatric patients and could possibly induce suicidal behavior. GlaxoSmithKline settled for \$2.5 million out of court (Kagle, 2008).

³ These journals include *European Journal of Work and Organizational Psychology*, *Group and Organizational Management*, *International Journal of Selection and Assessment*, *Journal of Business and Psychology*, *Journal of Managerial Psychology*, *Journal of Personnel Psychology*, *Journal of Vocational Behavior*, *Leadership Quarterly*, *Organizational Research Methods*, *Work, Aging and Retirement*. For more details, see *Journal of Business and Psychology* (2016).

TABLE 2
Topics for Doctoral Methodology Courses

Excessive simplification		
Consequences of dichotomizing findings.	Nonorthogonal projection of multidimensional spaces onto 2 or 3 dimensional subspaces. Coefficient changes when dropping correlated variables.	
Excessive complication		
Differences between analysis of history and predictions about the future.	Which properties of a time series are likely to extrapolate to future periods?	Green & Armstrong (2015)
Differences between analysis of a specific sample and generalizations to possible alternative samples.	Which properties of a sample are likely to generalize to other samples from similar populations? Value of parsimony. Ockham's Hill.	Armstrong & Green (2017) Cumming (2011) Gauch (2006)
Reasoning before versus after data analysis		
Differences between deduction, induction, abduction, and retroduction.	Analysis of "black boxes." Creative uses of statistical analyses. Post hoc data analysis.	Ashby (1956) Bunge (1963) Tukey (1977) Folger & Stein (2016) Hoaglin, Mosteller, & Tukey (1983) Hodgkinson & Starkey (2012) Locke (2007) Selvin & Stuart (1966) Silberzahn & Uhlmann (2015) Woo, O'Boyle, & Spector (2016) Kerr (1998) Simmons et al. (2011)
Consequences of covert HARKing and p-Hacking for evaluations of deductive hypotheses.	Dependence of "statistical significance" on prior hypotheses deduced before data analysis.	
Randomness of data		
Differences between analysis of a random sample and analysis of a subpopulation.	Dependence of statistical inferences on sample randomness and sample sizes. Finite-population correction for variance of the sample mean. How to analyze nonrandom data and subpopulations.	Cochran (1977) Knaub (2008)
Sample size and outliers.	Comparison with unit-weighted regression. Importance of large random samples to mitigate outliers. Robust regression and analysis of variance.	Bobko, Roth, & Buster (2007) Einhorn & Hogarth (1975) LeBel, Campbell, & Loving (in press) Rousseeuw & Leroy (1987)

workers in a specific factory; researchers might even describe observations based on such data as being "statistically significant." Since it is extremely unlikely that this factory hires from the entire population of the world or chooses workers by drawing random numbers, the data describe a complete subpopulation and so significance tests or *p* values are not even relevant. There seem to be many opportunities for methodology teachers to enhance students'

knowledge of alternative ways to analyze data and to use statistical analyses more effectively.

"There seem to be many opportunities for methodology teachers to enhance students' knowledge of alternative ways to analyze data and to use statistical analyses more effectively."

How Individual Researchers Can Help

Dishonesty is undoubtedly not characteristic of management researchers, who are, with very few exceptions, honorable people who believe in the goals of their research and want to make useful contributions to knowledge. Yet, studies have shown that many management researchers engage in practices that undermine the validity of their research, and some researchers do this in the belief that the behavior is correct. The inconsistency between values and behavior appears to be primarily a consequence of a social environment that has gradually grown more distorted over decades. Employers and mass-circulation periodicals reward "statistically significant" research and conformity to social norms. Concerns for publishability induce researchers to imitate the articles they see in journals. Journals publish what is submitted to them. Faculty research seminars show new entrants how research reports ought to look. The long-term result has been a drift away from excellent practices and toward deceptive ones.

One of the most tragic consequences of covert practices became visible while this article was being revised. A well-known researcher who bears the title "distinguished professor" told a trusted friend: "I have become increasingly concerned that due to *p*-hacking in many fields, we can't be sure if reported results are little more than Type 1 errors, even if they are replicated. It is becoming increasingly difficult to know what is and isn't real."

Every researcher should have a very strong interest in countering practices that produce such worries. Imagine that you produced many works of research that upheld high methodological standards, and you never intentionally misrepresented your practices or findings. But, you did rely on the articles in management journals for input to both teaching and research, and you published reviews of these articles. Then after 50 years or so, you now realize that half of what you have read was false. A first reaction is to reassure yourself that you can trust the studies you made and reported honestly, but then you realize that almost all of your reported findings depended on your calculations of *p* values, which are unreliable indicators of the likelihood of reproducing findings in new samples.

Researchers control research practices, data collection, and what and how they report. The ultimate quality of research articles hinges on researchers' being proud enough of their behaviors

that they can talk about them openly. Little lies are not necessary.

"The ultimate quality of research articles hinges on researchers' being proud enough of their behaviors that they can talk about them openly."

Repelled and embarrassed by the dishonest research practices that he thought were pervasive in quantitative research, Candide ceased doing such research even though he saw that a very high percentage of published studies entailed quantitative methods. Yet, Candide continued to be successful: He continued to publish in highly prestigious journals and to receive awards for excellent research. His continued success was partly a result of his activism. He did not assume that editors and reviewers would understand or appreciate qualitative methods, so his articles included rationales for his methodological choices. He developed inductive and abductive inferences as central themes in the abstracts and conclusions of his articles. Candide discovered that not all editors and reviewers demand conformity to ritualistic patterns, and some are as open-minded and curious as he is. Candide has also devoted some of his time to advocating and supporting change in quantitative methodologies by engaging in personal discussions, symposia, and workshops. He has discovered multiple ways to support needed methodological change.

All researchers should strive to maximize their contributions. Contribution depends on more than just getting articles published and often reveals itself only in hindsight after a substantial amount of evidence has accumulated. Newton and Darwin delayed publication of their brilliant works for years. Dressing up, streamlining, and cutting corners might help to get articles published but will damage others' ability to correctly interpret and build upon the reported findings. Publishing disconfirming findings, overt use of abductive reasoning, and iterative model development and comparisons promise to substantially enhance the quality of management research. Scientific progress hinges on motivating researchers not just to publish articles, but also to contribute to the accumulation of knowledge across studies with the ultimate goal of positive impact on management practice.

“Publishing disconfirming findings, overt use of abductive reasoning, and iterative model development and comparisons promise to substantially enhance the quality of management research.”

REFERENCES

- Academy of Management. 2006. Code of Ethics. Retrieved from: http://aom.org/uploadedFiles/About_AOM/Governance/AOM_Code_of_Ethics.pdf, Accessed February 15, 2017.
- American Psychological Association. 2010. *Publication manual of the American Psychological Association*, (6th ed.). Washington, DC: American Psychological Association.
- American Statistical Association Board of Directors. 2016. ASA statement on statistical significance and p-values. In Ronald L. Wasserstein, (Ed.), *The American Statistician*, 70(2): 131–133.
- Anonymous. 2015. The case of the hypothesis that never was; Uncovering the deceptive use of post hoc hypotheses. *Journal of Management Inquiry*, 24: 214–216.
- AOM Ethics Education Committee. 2011. Ethics video series: Reporting research. Retrieved from: <https://www.youtube.com/watch?v=fbFqLvuoSBM&feature=related>, Accessed January 30, 2016.
- Armstrong, J. S., & Green, K. C. 2016. Guidance for science: Evidence and checklists. Retrieved from: http://repository.upenn.edu/marketing_papers/181, Accessed February 16, 2017.
- Aschwendan, C. 2015. Science isn't broken: It's just a hell of a lot harder than we give it credit for. *FiveThirtyEight*, Retrieved from: <http://fivethirtyeight.com/features/science-isnt-broken>, Accessed August 21, 2015.
- Ashby, W. R. 1956. *The black box. An introduction to cybernetics*: Chapter 6, 86–117. London: Chapman & Hall.
- Bakker, M., & Wicherts, J. M. 2011. The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43(3): 666–678.
- Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A. A., & Adkins, C. L. 2016a. Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, 42(1): 5–20.
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. 2016b. Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, 31(3): 323–338.
- Bedeian, A. G. 2003. The manuscript review process: The proper roles of authors, referees, and editors. *Journal of Management Inquiry*, 12(4): 331–338.
- Bedeian, A. G., Sturman, M. C., & Streiner, D. L. 2009. Decimal dust, significant digits, and the search for stars. *Organizational Research Methods*, 12(4): 687–694.
- 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(4): 715–725.
- Bem, D. J. 2004. Writing the empirical journal article. In J. M. Darley, M. P. Zanna, & H. L. Roediger, III, (Eds.), *The compleat academic: A practical guide for the beginning social scientist*, (2nd ed.): 171–201. Washington, DC: American Psychological Association.
- Bergh, D. D., Sharp, B., Aguinis, H., & Li, M. 2015. The looming replication crisis in strategic management. Presentation at the Annual International Conference of the Strategic Management Society, Denver, CO.
- Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. 2016. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, 37(2): 257–261.
- Bhattacharjee, Y. 2013. The mind of a con man. *The New York Times* (August 23). Retrieved from: <http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?login=email&pagewanted=all&r=0>, Accessed June 6, 2016.
- Biemann, T. 2013. What if we were Texas sharpshooters? Predictor reporting bias in regression analysis. *Organizational Research Methods*, 16: 335–363.
- Bobko, P., Roth, P. L., & Buster, M. A. 2007. The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, 10: 689–709.
- Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. 2016. HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*, 69: 709–750.
- Bunge, M. 1963. A general black box theory. *Philosophy of Science*, 30(4): 346–358.
- Byington, E. K., & Felps, W. 2017. Solutions to the credibility crisis in management science. *Academy of Management Learning & Education*, 16(1): 142–162.
- Chang, A. C., & Li, P. 2015. Is economics research replicable? Sixty published papers from thirteen journals say 'usually not'. *Finance and Economics Discussion Series* 2015-083. Washington, DC: Board of Governors of the Federal Reserve System.
- Cochran, W. G. 1977. *Sampling techniques*, (3rd ed.). New York: Wiley.
- Cumming, G. 2011. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
- Einhorn, H. J., & Hogarth, R. M. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13: 171–192.
- Fanelli, D. 2013. Redefine misconduct as distorted reporting. *Nature*, 494: 149.
- Fiedler, K., & Schwarz, N. 2016. Questionable research practices revisited. *Social Psychological & Personality Science*, 7(1): 45–52.
- Fidler, F. 2005. *From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology*. Doctoral dissertation, University of Melbourne. Downloadable at <http://www.latrobe.edu.au/psy/staff/fidler.html>
- Folger, R., & Stein, C. M. 2016. Abduction 101: Reasoning processes to aid discovery. *Human Resources Management Review*.

- Retrieved from: <http://dx.doi.org/10.1016/j.hrmr.2016.08.007>. Available online August 30, 2016.
- Francis, G., Tanzman, J., & Matthews, W. J. 2014. Excess success for psychology articles in the journal *Science*. *PLoS One*, 9(12): e114255.
- Gauch, H. G., Jr., 2006. Winning the accuracy game. *American Scientist*, 94: 135–143.
- Goldfarb, B., & King, A. A. 2015. Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, 37: 167–176.
- Green, K. C., & Armstrong, J. S. 2015. Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68: 1678–1685.
- Greenwald, A. G. 1975. Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82: 1–20.
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.) 1983. *Understanding robust and exploratory data analysis*. New York: Wiley.
- Hodgkinson, G. P., & Starkey, K. 2012. Extending the foundations and reach of design science: Further reflections on the role of critical realism. *British Journal of Management*, 23(4): 605–610.
- Honig, B., & Bedi, A. 2012. The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education*, 11(1): 101–123.
- 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., & Lindsay, R. M. 2013. The significant difference paradigm promotes bad science. *Journal of Business Research*, 66: 1393–1397.
- Hubbard, R. 2015. *Corrupt research: The case for reconceptualizing empirical management and social science*. Los Angeles: SAGE Publications.
- Jensen, N. M. 2015. Data access and replication policies in management. Retrieved from: <http://www.natemjensen.com/2015/11/10/data-access-and-replication-policies-in-management/>; Accessed June 6, 2016.
- 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.
- Journal of Business and Psychology*. 2016. JBP Special initiative hybrid registered reports resource site. Retrieved from: <https://jbp.uncc.edu/>, Accessed January 30, 2016.
- Kagle, A. 2008. Driven to settle: Eliot Spitzer v. GlaxoSmithKline and undisclosed clinical trials data regarding Paxil. (August 18, 2008). Retrieved from: <http://web.law.columbia.edu/sites/default/files/microsites/career-services/Driven%20to%20Settle.pdf>
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. 2012. Publication bias in the organizational sciences. *Organizational Research Methods*, 15(4): 624–662.
- Kepes, S., Bennett, A. A., & McDaniel, M. A. 2014. Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education*, 13(3): 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. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3): 196–217.
- Knaub, J. 2008. Finite population correction (FPC) factor. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods*: 284–286. Thousand Oaks, CA: SAGE.
- 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 Psychology: Perspectives on Science and Practice*, 6(3): 299–302.
- LeBel, E. P., Campbell, L., & Loving, T. J. In press. Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology*. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2616384, Accessed: February 3, 2017.
- Locke, E. A. 2007. The case for inductive theory building. *Journal of Management*, 33: 867–890.
- Lovell, M. C. 1983. Data mining. *The Review of Economics and Statistics*, 65: 1–12.
- Mazzola, J. J., & Deuling, J. K. 2013. Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I–O journal articles. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 6: 279–284.
- McShane, B. B., & Gal, D. 2016. Blinding us to the obvious? The effect of statistical training on the evaluation of evidence. *Management Science*, 62(6): 1707–1718.
- McWilliams, A., Siegel, D., & Teoh, S. H. 1999. Issues in the use of the event study methodology: A critical analysis of corporate social responsibility studies. *Organizational Research Methods*, 2: 340–365.
- Merton, R. K. 1973. The normative structure of science. In N. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations*: 267–278. Chicago, IL: University of Chicago Press.
- Miller, A. N., Taylor, S. G., & Bedeian, A. G. 2011. Publish or perish: Academic life as management faculty live it. *Career Development International*, 16(5): 422–445.
- O’Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. 2014. The Chrysalis Effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 43(2): 376–399.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science*, 349(6251): 943, aac4716.
- Orlitzky, M. 2012. How can significance tests be deinstitutionalized? *Organizational Research Methods*, 15(2): 199–228.
- Peach, J. T., & Webb, J. L. 1983. Randomly specified macroeconomic models: Some implications for model selection. *Journal of Economic Issues*, 17(3): 697–720.

- Pfeffer, J. 1993. Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, 18(4): 599–620.
- Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86: 638–641.
- Rousseeuw, P. J., & Leroy, A. M. 1987. *Robust regression and outlier detection*. New York: Wiley.
- Schmidt, F. L., & Hunter, J. E. 2014. *Methods of meta-analysis: Correcting error and bias in research findings*. Newberry Park, CA: SAGE.
- Schwab, A., Abrahamson, E., Fidler, F., & Starbuck, W. H. 2011. Researchers should make thoughtful assessments instead of null-hypothesis significance tests. *Organization Science*, 22(4): 1105–1120.
- Schwab, A., & Starbuck, W. H. 2016. Collegial “nests” can foster critical thinking, innovative ideas, and scientific progress. *Strategic Organization*, 14(2): 167–177.
- Selvin, H. C., & Stuart, A. 1966. Data-dredging procedures in survey analysis. *The American Statistician*, 20(3): 20–23.
- Simonsohn, U. 2016. Evaluating replications: 40% full ≠ 60% empty. Retrieved from: DataColada.org/47, Accessed March 03, 2016.
- Sijtsma, K., Veldkamp, C. L., & Wicherts, J. M. 2015. Improving the conduct and reporting of statistical analysis in psychology. *Psychometrika*, 81(1): 33–38.
- Silberzahn, R., & Uhlmann, E. L. 2015. Crowdsourced research: Many hands make tight work. *Nature*, 526(7572): 189–191, <http://www.nature.com/news/crowdsourced-research-many-hands-make-tight-work-1.18508>.
- Siler, K., & Strang, D. 2016. Peer review and scholarly originality: Let 1,000 flowers bloom, but don't step on any. *Science, Technology and Human Values*, online before print, DOI 10.1177/0162243916656919.
- 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.
- Starbuck, W. H. 2016a. 60th anniversary essay: How journals could improve research practices in social science. *Administrative Science Quarterly*, DOI: 0001839216629644.
- Starbuck, W. H. 2016b. Squeezing lemons to make fresh lemonade: How to extract useful value from peer reviews. In T. Clark, M. Wright, & D. Ketchen (Eds.), *How to get published in the best management journals*: 85–108 Cheltenham: Edward Elgar.
- Starbuck, W. H., & Dutton, J. M. 1971. The history of simulation models. In J. M. Dutton & W. H. Starbuck (Eds.), *Computer simulation of human behavior*: 9–102. New York: Wiley.
- Sterling, T. D. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, 54: 30–34.
- Tukey, J. W. 1977. *Exploratory data analysis*. Reading, CA: Addison-Wesley.
- Wasserstein, R. L., & Lazar, N. A. 2016. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2): 129–131.
- Webster, J., & Starbuck, W. H. 1988. Theory building in industrial and organizational psychology. *International Review of Industrial and Organizational Psychology*, 3(1): 93–138.
- Wigboldus, D. H., & Dotsch, R. 2016. Encourage playing with data and discourage questionable reporting practices. *Psychometrika*, 81: 27–32.
- Woo, S. E., O'Boyle, E., & Spector, P. E. 2016. Best practices in developing, conducting, and evaluating inductive research. *Human Resources Management Review*, Retrieved from: <http://dx.doi.org/10.1016/j.hrmr.2016.08.004>, Available online August 22, 2016.

Andreas Schwab (aschwab@iastate.edu) is an associate professor and dean's fellow in the College of Business at Iowa State University. He is a contributing editor for Research Methodology and Statistics at *Entrepreneurship, Theory & Practice*. He studies organizational learning in project ventures and is engaged in several initiatives to advance methodological practices in management science.

William H. Starbuck (starbuck@uoregon.edu) is a visiting professor at University of Oregon and professor emeritus at New York University. He edited *Administrative Science Quarterly*, chaired the screening committee for Fulbright awards in business management, and was president of the Academy of Management. His current research interests are research methodology, innovation, and societal trends.