

The Gray Zone: Questionable Research Practices in the Business School

NICK BUTLER

Stockholm University, Sweden

HELEN DELANEY

University of Auckland, New Zealand

SVERRE SPOELSTRA

Lund University, Sweden

In recent years, the awareness of academic misconduct has increased due to high-profile scandals involving prominent researchers and a spike in journal retractions. But such examples of fabrication, falsification, and plagiarism (FFP) serve to obscure the less flagrant, more subtle cases of possible misconduct: what some have called “questionable research practices” (QRPs). Where FFP is seen as inherently negative, QRPs fall into an ethical “gray zone” between permissible and impermissible. We draw on semistructured interviews with business school scholars to explore the occurrence of QRPs. Prevalent QRPs include playing with numbers, playing with models, and playing with hypotheses. Scholars explain the existence of QRPs in three ways: the inadequate training of researchers, the pressures and incentives to publish in certain outlets, and the demands and expectations of journal editors and reviewers. We argue that a paradox is at work here: To live up to the positivist image of “pure science” that appears in academic journals, researchers may find themselves—ironically—transgressing this very ideal. Ultimately, this challenges the individualistic account of academic misconduct by drawing attention to the role played by institutional actors, such as academic journals, in encouraging forms of QRPs.

In recent years, the awareness of academic misconduct has increased due to high-profile scandals involving prominent researchers and a spike in journal retractions (Fang, Steen, & Casadevall, 2013; Marcus & Oransky, 2015; *The New York Times* Editorial, 2015; Van Noorden, 2011). Recall, for example, the case of disgraced social psychologist Diederik Stapel, who fabricated entire data sets for nonexistent experiments over the course of nearly 2 decades (Bhattacharjee, 2013). When his fraud was eventually brought to light in 2011, 58 of his academic journal articles were retracted and struck

from the scientific record (Retraction Watch, 2015). Predictably, this resulted in an extended period of soul searching within the field of social psychology and a wider discussion about how to strengthen the mechanisms for detecting and preventing misconduct in academic research (Budd, 2013; Stroebe, Postmes, & Spears, 2012).

But such examples of outright fraud arguably serve to obscure the less flagrant, more subtle cases of potential misconduct, or what some call “questionable research practices” (QRPs) in contrast to fabrication, falsification, and plagiarism (FFP; Fanelli, 2012). Although by definition there is no agreement about what type of research practices fall into the category of “questionable,” they commonly involve

This research is funded by the Swedish Research Council (project no. 2015-01100) and Handelsbanken’s Jan Wallander and Tom Hedelius Research Foundation (project no. P2015-0106:1).

misrepresentation, inaccuracy, or bias (Steneck, 2006). QRPs often include, for example, changing hypotheses after the fact, falsely attributing authorship, omitting outliers, and salami slicing data. Wherever one chooses to draw the boundary, FFP are seen as inherently negative, whereas QRPs fall into an ethical "gray zone" (Lynöe, Jacobsson, & Lundgren, 1999: 501) between acceptable and unacceptable. This makes QRPs a fruitful starting point to discuss research ethics within a field of academic study. Prior research on academic misconduct, with some notable exceptions (e.g., De Vries, Anderson, & Martinson, 2006; Wenger, Korenman, Berk, & Berry, 1997), has tended to adopt a quantitative approach to studying the incidence and prevalence of FFP and QRPs (see, e.g., Bedeian, Taylor, & Miller, 2010; Honig & Bedi, 2012; Fanelli, 2009; List, Bailey, Euzeit, & Martin, 2001). Such studies are valuable for showing the rate of perceived misconduct in academia. However, qualitative data offers a rich interpretive perspective on where—and why—individual scholars draw the line between permissible and impermissible research practices.

Our work here draws on extensive semistructured interviews with business school scholars to explore the occurrence of QRPs. We begin by reviewing research on academic misconduct and questionable research practices within management and organization studies and further afield. After describing our qualitative methods, we present our empirical material. We show that QRPs in the business school include playing with numbers, playing with models, and playing with hypotheses. Our respondents explain the existence of such QRPs in three ways: the inadequate training of researchers, the pressures and incentives to publish in certain outlets, and the demands and expectations of journal editors and reviewers. In the Discussion section, we identify a paradox: To live up to the positivist image of "pure science" that appears in academic journals, researchers may find themselves—ironically—transgressing this very ideal. Ultimately, this challenges the individualistic account of academic misconduct by drawing attention to the role played by institutional actors, such as academic journals, in encouraging—implicitly or explicitly—forms of QRPs.

ACADEMIC MISCONDUCT AND QUESTIONABLE RESEARCH PRACTICES

The question of academic misconduct is becoming increasingly prevalent in management and organization studies. This is indicated by recent editorials that discuss recent retractions involving the

work of prominent management and organization scholars (Retraction Watch, 2014a, 2014b). For example, present and former editors of *The Leadership Quarterly* comment that, while "retraction decisions do not imply scientific misconduct" (Atwater, Mumford, Schriesheim, & Yammarino, 2014: 1175), they are on the rise because "authors seek publication in high visibility journals by any means, fair or foul" (2014: 1179). The implication is that some articles may not meet acceptable standards of scholarship, yet somehow manage to slip through the net of peer review, whether due to honest mistakes or out-and-out fraud. It is worth noting that articles are far more likely to be retracted because of intentional misconduct than innocent errors, at least in biomedical and life sciences (Fang, Steen, & Casadevall, 2013). Another editorial, in the *Journal of Management Studies*, notes the growing concern with "unethical behavior in publishing," which is seen to result from "heightened competition" in academia and the "pressure to publish" in highly rated outlets (Harley, Faems, & Corbett, 2014: 1361–1362). Proposed solutions to counteract misconduct and reduce the number of retractions include implementing more stringent codes of conduct for authors and tightening editorial and peer-review processes in journals (see, e.g., Atwater et al., 2014; Chen, 2011; Harley, Faems, & Corbett, 2014; Schminke, 2009).

"The implication is that some articles may not meet acceptable standards of scholarship, yet somehow manage to slip through the net of peer review, whether due to honest mistakes or out-and-out fraud."

The problem with this approach is that what counts as unethical behavior is not always self-evident, particularly when it comes to borderline cases of misconduct (Chubin, 1985). Of crucial importance is a distinction between proscribed "black" practices and potentially inappropriate "gray" ones (Fanelli, 2012), commonly articulated in terms of fabrication, falsification, and plagiarism (FFP) on the one hand and "questionable research practices" (QRPs) on the other. QRPs are in an ethical "gray zone" (Lynöe, Jacobsson, & Lundgren, 1999: 501) because they fall in between FFP on the one hand and scholarly best practice or the "responsible conduct of research" on the other (Steneck, 2006).

The notion of “questionable research practices” was formalized in 1992 in a report published by a committee comprised of the National Academy of Sciences, the Institute of Medicine, and the National Academy of Engineering (Panel on Scientific Responsibility and the Conduct of Research, 1992; Resnik, 2003). By developing the concept of QRP, the committee sought to distinguish outright fraud from other practices that “erode confidence in the integrity of the research process, violate traditions associated with science, waste time and resources, and weaken the education of new scientists” (Panel on Scientific Responsibility and the Conduct of Research, 1992: 28). Although the report (1992: 28) mentions QRPs, such as failing to keep adequate research records, withholding data that support published research, inaccurately attributing authorship, and using statistics and other methods to artificially enhance the significance of research findings, the list of gray practices has since grown to encompass post hoc hypothesizing, redundant publication, salami slicing data, selective reporting of results, and inadequate or inaccurate citation, among others (Steneck, 2006).

In 2000, the U.S. federal government adopted a definition of academic misconduct as FFP, excluding QRPs from its remit (Office of Science and Technology Policy, 2000; Resnik, Neal, Raymond, & Kissling, 2015). This is a shame because, as the Panel on Scientific Responsibility and the Conduct of Research acknowledged (1992: 29), there is potential overlap between black, gray, and white research practices. In other words, the distinction between “obvious misconduct” and “normal scientific practice” is not entirely clear-cut (Lynøe, Jacobsson, & Lundgren, 1999: 501). For example, some scientists acknowledge that the line between legitimately “cleaning” data and inappropriately “cooking” data are inherently ambiguous because a researcher’s scientific judgment is shaped by prior experience—that is, what they expect data to tell them about a phenomenon (De Vries, Anderson, & Martinson, 2006: 46). QRPs thus underscore the fact there is no “demarcation principle” between acceptable and unacceptable research practices (Lynøe, Jacobsson, & Lundgren, 1999: 501). The stakes of studying QRPs now become clear: If our aim is to promote research integrity and research ethics, rather than simply to expose and punish wrongdoers for their flagrant transgressions, then we must take the gray zone into full consideration (Fanelli, 2012).

Studies have shown that questionable research practices are far more widespread than cases of

fabrication, falsification, and plagiarism (Fanelli, 2009; John, Loewenstein, & Prelec, 2012; Martinson, Anderson, & de Vries, 2005), although we should be careful not to overstate the case (Fiedler & Schwarz, 2015). In his meta-analysis of surveys on scientific misconduct, Fanelli (2009: 8) states that while 2% of scientists admit to FFP, almost a third admit to engaging in QRPs. One possible explanation for the prevalence of QRPs may be due—ironically—to the increasing awareness of FFP and other forms of overt misconduct. Fanelli (2013a: 5) writes:

Rather like professional athletes, who strive to maximize performance-enhancing practices within the allowed limits, scientists might be getting better at “pushing” their findings in the desired direction and stopping right before the “misconduct threshold” or at avoiding getting caught if they trespass it.

In other words, scholars may seek to avoid institutional sanctions by engaging in research practices that artificially inflate the significance of their findings in some way, yet fall short of outright misconduct.

In management and organization studies, questionable research practices are no less pervasive. As Bedeian, Taylor, and Miller (2010) show, the vast majority of respondents to a survey of 384 academics reported knowing of faculty who engage in both FFP and QRPs. Some fraudulent activities, such as fabricating data, were reported by 26.8% of respondents, while some QRPs, such as inaccurately attributing authorship and hypothesizing after the fact were reported by 78.9% and 91.9% of respondents, respectively (2010: 716). Of course, reporting on the research practices of others is not the same as reporting on one’s own research practices. Even so, the survey minimally suggests that QRPs in management and organization studies are “by and large, even more common than outright misconduct” (2010: 719; see also Lenz, 2014; for a similar study in economics, see Necker, 2014). Certainly, this is the conclusion drawn in recent exposés of post hoc hypothesizing in the field, both quantitative (O’Boyle, Banks, & Gonzalez-Mulé, 2017) and anecdotal (Anonymous, 2015).

In light of this, a neglect of QRPs might lead to misapprehensions about the nature of academic research. As De Vries, Anderson, and Martinson (2006: 44) acknowledge, an overemphasis on outright fraud—particularly headline-grabbing cases of

retraction—may inadvertently “give the appearance of confirming the integrity of science: wrongdoers are caught and disciplined, assuring the public that the bad apples of science cannot long survive.” It is therefore important to focus on QRPs to bring to light questionable—although not strictly outlawed—research practices that tend to escape critical attention due to their very “ordinariness.” Indeed, De Vries and colleagues go on to suggest that the study of so-called “normal misbehavior,” following Durkheim, plays a useful role in academic research. This is because a focus on QRPs serves to highlight the “pinch points” that characterize scientific inquiry, such as ambiguities around methods and procedures (2006: 50). By ignoring QRPs, we risk overlooking a set of widespread, yet partially hidden, practices that shape the rules of engagement in academic research.

Perpetrators of fraud may reinforce an individualistic view of academic misconduct, as we find in Diederik Stapel’s admission that “I will have to dig deeply to find out myself, why all of this happened, and what motivated me to do this” (cited in Stroebe, Postmes, & Spears, 2012: 672). It is unlikely, however, that introspection will get us very far. Indeed, individualistic explanations for misconduct serve to obscure other, more systemic issues around QRPs. For instance, the recent retraction of Michael LeCour’s discredited study of political canvassing from *Science* may have made international headlines (Carey, 2015; Gambino & Devlin, 2015), but what is little mentioned is the role played by the other author on the paper, Prof. Donald Green. As the editorial expression of concern (published prior to the retraction) states, “Green requested that *Science* retract the paper because of the unavailability of raw data and other irregularities that have emerged in the published paper” (McNutt, 2015: 1100). This begs the question: Should a coauthor (honorary or otherwise) put his or her name to a piece of work without first having access to the raw data on which the findings are based? It is below the threshold of academic misconduct, yet arguably falls within the remit of “questionable.” On this point, we concur with Sovacool’s (2008: 280) suggestion that singling out “a few individual violators ... creates the illusion of solidarity among the scientific community ... And by isolating a few behaviours as corrupt, it stamps all others as blameless.” This highlights the need to consider not only the actions of individual fraudsters, but also the structures in which their activities are situated.

“By ignoring QRPs, we risk overlooking a set of widespread, yet partially hidden, practices that shape the rules of engagement in academic research.”

In particular, it is important to consider the decisive role played by prestigious academic journals in creating an environment in which QRPs flourish (Bedeian, 2003; Davis, 2014; Macdonald, 2015). For example, journal editors may employ various “tricks of the trade” to artificially inflate the impact factor and boost the ranking of their journals, aided and abetted by for-profit publishing companies (Butler & Spoelstra, 2015; Harvie, Lightfoot, Lilley, & Weir, 2012, 2013; Macdonald, 2015; Martin, 2013; Schekman, 2013). By skewing the production of knowledge in academia, journal editors encourage authors to “play the game,” and thus, potentially engage in QRPs, to increase their chance of publication. As Kerr (1998: 205) notes, editors and reviewers may “sanction or even insist upon HARKing”—hypothesizing after the results are known—as well as other QRPs (Kepes & McDaniel, 2013; O’Boyle, Banks, & Gonzalez-Mulé, 2017). This is a far cry from the atomistic approach to misconduct found in recent editorials in the field, which deflect attention away from their own complicity in QRPs (see, e.g., Atwater et al., 2014; Chen, 2011; Harley, Faems, & Corbett, 2014; Schminke, 2009).

It is crucial, therefore, to understand the context in which QRPs take place. As Martinson, Anderson, and de Vries (2005: 738) write: “Missing from current analyses of scientific integrity is a consideration of the wider research environment, including institutional and systemic structures.” The remainder of our article addresses this lacuna by studying QRPs in the business school, taking a qualitative approach to examine the incidence and explanations for “normal misbehavior” among leadership scholars.

METHOD

This article is part of a broader project about the conditions of knowledge production in the business school (Butler, Delaney, & Spoelstra, 2015). Taking our cue from reflexive and critical research in the qualitative tradition (Alvesson & Skoldberg, 2009; Prasad, 2005), our focus is on leadership scholars (i.e., full-time academic researchers who study leaders and leadership) and how they understand and navigate the competing pressures and tensions

shaping academic labor—including, but not limited to, questionable research practices. Leadership studies is comprised of scholars from a variety of paradigmatic backgrounds, including positivism, interpretivism, and critical management studies. As such, we conducted in-depth, semistructured interviews with scholars from each of these communities. We selected interviewees according to three (sometimes overlapping) criteria: membership of editorial boards of leadership journals, such as *The Leadership Quarterly* and *Leadership*; membership of leadership centers and university departments that are associated with leadership studies; and prominent contributions to leadership studies in academic journals or scholarly monographs. We then contacted respondents through email and conducted interviews face-to-face or, in 10 cases, over Skype.

We began with seven pilot interviews in July 2011. This allowed us to further refine our interview questions for the main period of data collection between October 2012 and August 2014, which involved all authors of this article. In total, we conducted interviews with 72 leadership scholars based in North America, the United Kingdom, Europe, and Australasia, of which 43 are male and 29 are female. Twenty-nine of our interviewees employ positivist, quantitative methods in their research, 26 use qualitative methods from an interpretivist perspective, and 17 adopt a critical stance toward leadership using social theory or philosophy (although the distinction between the latter two groups is not clear-cut). Overall, we interviewed 43 professors, 21 associate professors, senior lecturers, and readers, and 8 assistant professors and lecturers. We ceased conducting new interviews when we reached a point of data saturation, that is, when no novel information emerged from the groups of scholars we interviewed in relation to our key themes (Corbin & Strauss, 2008). The longest interview was 2 hours, the shortest was 30 minutes, and the average was 1 hour. All interviews were recorded and fully transcribed, and the names of our participants have been changed.

From the outset, we sought to examine how ideas of “rigor” inform the working lives of the research participants. In particular, we were interested in what constitutes “rigorous” research among leadership scholars and the extent to which their own work lives up to these ideals. We soon realized that many scholars were questioning the robustness of leadership studies as a scientific pursuit, and a few

scholars alluded to the potential for misconduct in the field. In February 2014, our attention was drawn to several articles published in *The Leadership Quarterly* that were going to be retracted (which they were in December 2014) and that other highly ranked management journals were investigating similar cases. Although no intentional wrong-doing is inferred from these retractions, they catalyzed our interest in questionable research practices. Hence, in our interviews from February 2014 onward (encompassing 29 leadership scholars), we included more specific questions about where our respondents draw the line between appropriate and inappropriate research practices. We broached this topic first by asking our respondents how they distinguish between “good” and “bad” science in leadership studies and how far they feel their research lives up to this ideal, before asking how they make sense of the LQ retractions. This opened up a space to discuss whether they had ever felt any temptation to engage in research practices they consider to be questionable. We posed these questions to positivist researchers as well as to interpretivist and critical researchers, with some variation in language to fit the norms of each paradigm. Although we focus on leadership studies, it is part of a wider field of management and organization studies that is also characterized by a strong positivist tradition. Therefore, our findings and recommendations plausibly apply to business school research beyond the field of leadership.

We focus here on the interviews with positivist leadership scholars because they, unlike interpretivist and critical scholars, consider themselves to be working within science “proper” (Atwater et al., 2014). Any potential transgression of responsible research practices—whether “black” or “gray”—will therefore be of prime importance to scholars in the field. This is not to say, of course, that interpretivist and critical research is free from FFP or QRPs. Indeed, two of the authors of this paper have previously researched academic “game-playing” in critical management studies, which includes establishing multiauthorship cartels, acceding to unreasonable requests from editors and reviewers, watering down critical discussion to increase the likelihood of publication in elite journals, and other questionable research practices (Butler & Spoelstra, 2012, 2014).

As scholars employed by business schools, we are familiar with many of the tensions and pressures that our interviewees discussed, such as the expectation to publish in certain outlets or the

demands and expectations of journal editors and reviewers. We are also familiar with the history, theories, and "camps" that comprise leadership scholarship: Two of us have contributed research to this field, so we are not entirely outside of the field we are investigating. Sharing this common ground facilitated greater access to interviewees and enabled us to ask more specific questions. However, as qualitative researchers, it was our first encounter with the intricacies of positivist research. This meant that we needed to learn the assumptions, language, and dynamics of this community, as well as understand the politics, divisions, and competing positions among positivist researchers. For this reason, it is possible that our status as outsiders to the positivist community may, at times, have hampered our fieldwork. For example, interviewees may have (intentionally or not) constructed certain interpretations that serve to "protect" their community from the critical gaze of an interloper. On the whole, however, we feel our status as outsiders enabled us to ask fruitfully naïve questions as we genuinely tried to understand the interviewees' mind-sets and the composition of their community. Indeed, many respondents seemed comfortable going into detail about "gray zone" practices with us precisely because we are *not* fellow positivist researchers.

In terms of data analysis, all authors read the transcripts of the interviews and collectively discussed emerging themes. In line with the principles of interpretivist research (Corbin & Strauss, 2008), we paid attention to the respondents' assumptions, meanings, and judgments, particularly in terms of "good" and "bad" science as well as "gray zone" practices. We were interested in any tensions or contradictions between their stated ideals of science and their day-to-day research practices. The first and second authors subsequently returned to the data set and, dividing the interviews into two groups, each researcher undertook a thorough rereading of their allocated transcripts. We organized excerpts of data into a number of main themes and, after several iterations and further discussion, we settled on two overarching themes that best capture the breadth and depth of data: (1) examples of QRPs, and (2) explanations for QRPs. These two themes form the structure of our Findings section. After we organized the raw data into these two themes (and associated subthemes), the two authors took responsibility for writing the first draft of a given subtheme. Following this, the other authors would check, nuance, clarify, and challenge the analysis,

at times returning to the transcripts to check the accuracy of data and to add more texture.

As we saw in the previous section, one common QRP is the misleading attribution of authorship to journal articles. For the sake of transparency, we feel it is important to describe the division of labor during our writing process. The first author took the lead on writing the introduction, literature, and Recommendation sections; the second author took the lead on writing the Method section; the first and second authors took a lead on writing the Findings section; and the third author took the lead on writing the Discussion and Conclusion sections. Once drafted, all authors were involved in further discussing, refining, and revising the paper for publication, contributing equally to crafting the overall focus and argument.

FINDINGS

Questionable Research Practices in the Business School

Many of our respondents hold a belief in leadership studies as a science. This expresses itself in a commitment to epistemological objectivity. As one respondent puts it: "Our job as scientists is to ... basically conduct research and let the chips fall where they do" (Hugo). There is an underlying assumption among our respondents that leadership is a phenomenon that can be known through the application of proper scientific methods, usually statistical in nature. Of crucial importance is the accurate presentation of methods, data, and results in academic articles:

You have to say here's what I did, here's how I did it, here's why I did it—boom! This is basic science ... If you don't say that, you can't replicate the observational conditions, right? (Cornelius)

Cornelius' appeal to objectivity and replication is in line with a positivist image of science, which holds that science can be clearly demarcated from non-science. For these respondents, leadership is a science in the same way as the natural sciences. Despite this plea for science "proper," following the hypothetico-deductive method, many respondents harbor doubts about some research practices that have become common in the business school. In the following, we explore how leadership scholars make sense of these questionable practices.

Playing With Numbers

The first QRP we identified involves the reporting of data. This is brought to the fore in terms of omitting outliers from statistical analyses. For example, Portia tells us about a former colleague who had published an article in a premier outlet:

I know that they left out 10 or 12 outliers in one of the data sets ... but they didn't communicate this to the journal ... Without these 10 or 12 outliers, they could replicate the study, and then they had a high chance to get this publication ... They just left these people out, and they said, "But if we communicate it, it will be rejected." (Portia)

While some may see this as a clear-cut case of academic misconduct in terms of the "false reporting of results" (Renaldo), others suggest that it is a gray zone:

It's unfeasible to disclose every assumption that you made about the data set. What do you do about omitting outliers? What do people do, given a raw data set, how do they go about cleaning that? There are a million different things that go into that. What's appropriate and what's not appropriate? (Caleb)

For Caleb, the line between acceptable and unacceptable practices is not clearly demarcated, because one cannot account for every decision made in the process of data analysis. The implication is that the distinction between "cleaning" and "massaging" the data are extremely blurred. In light of this uncertainty, Jacob suggests that authors are often happy to allow editors and reviewers to draw this line themselves:

There has been some gamesmanship in terms of, "Oh look, if I do this, look what happens to my results," and not fully reporting what they've done, and letting the peer review process evaluate whether that was appropriate or not. (Jacob)

The problem, of course, is that editors and reviewers may not be fully informed about the extent of data manipulation involved in a given study, and consequently, highly ranked journals may find themselves publishing articles of dubious academic quality. As Portia notes, "I can just delete like 100 data points [and] you would never know it. How

would you know? How would anybody find out?" For such respondents, the recent retractions in *The Leadership Quarterly* and elsewhere are simply "the tip of the iceberg" in terms of the degree of data manipulation in the field:

All the data sets I have seen [are] never perfect, never. And then you see these perfect studies in these top journals and think, "It's not possible, it's just not possible." (Portia)

This view is reflected in Maybell's misgivings about a paper she was involved with early in her career that drew on data that, in her view, seemed to have been inappropriately manipulated:

The data set was given to me as a doctoral student and I looked at it ... and I couldn't get any results out of it. And so I went to my adviser and I said, "There's nothing here, there's nothing we can do with this." Well, fast forward a couple of months ... and my adviser got in touch with me and he said, "Look, somebody got in touch with us [about] a special issue ... I think what we'll do is, we'll take your data set and give it to this guy, he'll find results" ... I was like, how is he going to find results? What is he going to do that I didn't do in my analysis? And so I just wrote the theory part, and this other guy did the methods and the results, and that paper has been published and it has been widely cited, and every time I see that it's been cited, I shake my head and think, "How is that possible?" (Maybell)

This example touches on several questions about research practices in the business school, such as the incentive to publish in academic journals and the division of labor in the research process. It is striking that the paper was published in a reputable journal and has gone on to have considerable impact in this field, despite one of the authors expressing serious doubts about its credibility. This suggests that playing with numbers may be tolerated even by those who are aware of possible manipulation.

Playing With Models

The second QRP we identified is the manipulation of models that are used for measuring a leadership construct (e.g., "transformational leadership," "authentic leadership," "ethical leadership"). There are a number of ways models can be altered, such as how error is correlated, how fit is calculated, and

how the items of a model are grouped together. Jacob describes the latter of these “very common” but “highly inappropriate” practices in more detail:

You have this leadership construct [which] consists of dimension A, dimension B, dimension C, dimension D. And each one of these dimensions has items that indicate it ... These things are not necessarily interchangeable; they measure a different facet or component of this leadership skill. Then, as you move forward to research, you should treat them as separate predictors in your research design. But that eats up degrees of freedom, it calls for more participants in the research. So, what researchers have been doing is showing that these dimensions are highly correlated and then saying, “Well, a higher order construct explains these underlying dimensions, and so we’ll average them together and call them “X” leadership.” (Jacob)

In other words, researchers are collapsing into a broader group individual dimensions that in theory should remain separate (since they purport to measure a different aspect of leadership skill). Hugo explains this is a “very serious problem,” because scientific decisions should be based on established theory or statistical rules. However, in practice other forces come into play:

People often just use whatever produces the most favourable results in terms of demonstrating that their data should be taken, you know, seriously [...] The result is that when they estimate their models, they do things that theoretically make little sense, but that statistically produce what seems to be a better result. By “better,” I mean publishable. (Hugo)

Hence, researchers are “playing with models” (Diane) to produce results that “favour the authors” (Jacob) in terms of increasing the likelihood of publication. Although some researchers appear to hold to strict rules around how measurement should be performed in leadership studies, others argue that these rules are “not nearly as worked out as clearly” (Quentin) as they appear. Indeed, Quentin sees the ability to “change our models in lots of different ways” as part of “good exploratory research.” He admits that his team has “done lots of studies where we fool around and try and get measures to fit, we just don’t try and publish them.” Rather, Quentin uses this “exploratory research” to inspire new studies. For him, the problem is not “fooling

around” with models, but that “people push it too far” when they try to publish such research. But this raises the inevitable question: Just how far is too far?

Playing With Hypotheses

The third QRP we identified concerns the development and testing of hypotheses. Of particular concern is the “very gray area” (Abe) involved in constructing an hypothesis after the data have been collected and analyzed, a practice that is also known as HARKing (“hypothesizing after the results are known,” or, more concretely, presenting a post hoc hypothesis as if it were an a priori hypothesis). HARKing involves a degree of deception, because editors, reviewers, and readers will assume—unless the author tells them otherwise—that the study has followed the standard deductive method, which involves developing a hypothesis and then testing it. Jacob explains how research should ideally be conducted:

You’ve got a set of hypotheses which you’ve built out of the theoretical foundations for them, you’ve designed your research accordingly, you measure appropriately, and then you test things and come up with some idea of whether or not ... your model fits the data and whether the hypotheses in terms of relationships are there.

By contrast, he describes HARKing in the following way:

You’ve got a data set ... you’ve run some regressions or some structural equation modelings, you see what fits together, and then you hypothesize things after the results are known.

Respondents tend to be in agreement that HARKing is “fairly routine” (Joseph). However, there is little agreement about whether it is acceptable. For Jacob, HARKing is simply “bad science” and “obviously a major violation of research protocol and research ethics.” Lisbeth, on the other hand, views HARKing as “quite normal because you spend so much time and effort” collecting empirical material, and “fishing” within the data may lead to “really cool” insights that the researcher didn’t expect. Likewise, Quentin admits to “mining” his data “to come up with hypotheses.” Although he states that he does not usually publish the results of this kind of analysis, he will make an exception if he thinks the findings are particularly interesting or difficult to

replicate due to the unique conditions under which the data were collected. Similarly, Abe agrees that "some kind of jiggling" is permissible so long as the researcher is comfortable that it as a "reliable way of seeing the data," and they are not doing to it "get a publication or maybe save [their] tenure." Ultimately, Abe feels the review process should determine whether his method is reasonable, although Jacob highlights how difficult it is for reviewers to determine whether HARKing has taken place, given that researchers conceal such practices by claiming that they have followed a deductive method.

Explanations for Questionable Research Practices

Insufficient Methodological Training

Our respondents offer a number of explanations for why they and others engage in QRPs. The first relates to insufficient methodological training, which applies to both junior and senior scholars. Many senior scholars, now responsible for training junior scholars, apparently were not trained adequately in very sophisticated methods. Diane admits, "all this SEM [structural equation modelling] stuff came in after we went through school. Unless you're really a statistician, you wouldn't keep up." Wendell, who was trained in social psychology, bemoans "the lack of rigour" in a lot of the leadership graduate programs "where the level of statistics and research methods isn't much more than what an undergraduate psych major would get—maybe not even that much." As a result, doctoral students "aren't getting what they need to really be good scientists" (Hugo), and senior scholars are left to question "whether you're making the right decisions" in their methods (Jacob).

Researchers are also struggling to keep up with innovations in statistical software. Hugo elaborates:

One of the things that has happened with the advent of canned computer programs is that people can do very sophisticated analyses without really understanding the theoretical and statistical implications of what they're doing. And so it's, you know, point and click, and the result is that ... we have these enormously powerful methods being misapplied by people who really don't even understand that they're misapplying them.

Although some researchers try to stay "as informed as possible," many admit they lack the statistical

knowledge to do so (Jacob). The peer-review process is pitched as the solution to these inadequacies: "You're hoping that through friendly reviews and peer reviews that you have people who are informed enough to tell you that you are or are not making the right decisions with your analysis" (Jacob). Following this explanation, QRPs are seen as unintentional mistakes, which begs the question: "Who do you blame: the trainer or the trainee?" (Diane).

However, a number of our respondents suggest scholars may in fact be trained into some QRPs. As Diane describes: "It's all playing with models, and that's how people are trained ... [It's] like, 'Go play with the data and the models until you can get it to work right.'" Similarly, Portia reflects on her training as a PhD student during which her supervisor taught her that if "an item is not really working, then we just leave it out ... I'm not sure if it's really okay... why would I criticize my professor? I thought she knew how it works ... You just follow." This suggests that certain kinds of QRPs, such as playing with numbers or playing with models, may be part of the process of educational training for leadership scholars.

Pressures and Incentives to Publish

The second explanation of why researchers engage in QRPs is "the pressure to publish" (Abe). Again, this applies to both junior and senior scholars. For junior scholars, "it's difficult to get a job, you've got to have publications for a job" (Allegra), let alone "tenure and promotion" (Hugo). Therefore, some junior scholars work with "big name people" (Allegra) in the hope it may boost their chances of publication. As a result, they may be less inclined to "be whistleblowers" if they witness any QRPs (Allegra).

This same perceived pressure to publish also applies to senior academics, albeit with different dynamics. Full professors with tenure, for example, may be concerned about securing professional and material rewards gained through top-tier publications. For example, some business schools offer large financial incentives for publishing in premier outlets, which means that "[if] you haven't done anything lately, you can sort of lose status, lose resources [and] lose power" (Wendell). In this case, the "pressure" is largely self-imposed and motivated by personal gain. Clarence speaks hypothetically of the "seductive" nature of being a prolific leadership scholar, which he suggests may lead to forms of QRPs: "I can imagine why he [sic] would manipulate his data because he's flying around the world meeting interesting people, staying in cool

places, doing good things." Senior scholars may therefore engage in QRPs not because they are worried about their jobs, but because they may become "too greedy for fame" (Diane).

Demands and Expectations of Journals

The third explanation relates to the demands by editors and reviewers of top-ranked journals for studies to conform to particular expectations about what research should look like. For example, many respondents told us that leadership journals are becoming less interested in studies that test null hypotheses. Amanda elaborates:

If I do a project that doesn't support the [leadership] model or the theory, I might as well throw the results in the trash because nobody is going to publish it ... We were trained to look for disconfirming evidence; now disconfirming evidence is useless.

Some suggest that because leadership studies is inclined toward practical relevance, journals are keen to publish studies that produce confirmatory hypotheses:

Being practice oriented, being prescriptive in nature, then [translates] to us wanting to say, "This is what we know." The only way you think you might know it is to confirm the hypothesis. Because if you don't confirm it, all you know is you don't know. (Clarence)

This means that, in the pursuit of relevance, journals will have a bias toward positive rather than negative hypotheses. The problem is that during the review process "you end up having to craft a paper that the reviewers want you to write, not a paper that [you] actually wrote" (Abe). As a result, he continues, "it's kind of almost a post hoc fit of theory to justify the research." Clarence puts it succinctly: "The field is making us do this. They don't publish non-significant results." Lisbeth elaborates:

I'm not blaming the journals, or the reviewers or the editor, but they also make you do things that are actually in a gray zone. So for instance making you come up with an hypothesis that you didn't have before, and then test it and then report that as a confirmatory analysis in the paper. Which is actually not allowed.

It is interesting to note that Lisbeth does not blame journals, reviewers, and editors for encouraging QRPs such as HARKing, but perhaps this is because authors are equally complicit in such "gray zone" practices. As Lisbeth says: "It's important to have high impact papers, and if a journal asks you to do that ..." She doesn't need to finish the sentence, because the implication is already clear enough: Authors will be inclined to play along for the sake of securing highly ranked publications. Similarly, Jan tells us about being explicitly instructed by a premier outlet to "change the method" in line with the editor's own research agenda. Such coercive editorial practices may be condoned by authors because, as Jan acknowledges with a laugh, "I want to get this article published." Ironically, this serves to produce a corpus of knowledge in top-ranked journals where "all the data are perfect and everything works out fine" (Lisbeth), effacing the traces of QRPs in the canon of leadership studies.

Of course, the peer-review process can also help to improve a paper and strengthen its contribution. But it is striking that scholars are willing to outsource their judgment to editors and reviewers insofar as the line between acceptable and unacceptable behavior is concerned. As we saw, it often comes down to a question of "letting the peer review process evaluate whether that was appropriate or not" (Jacob). However, editors and reviewers may explicitly condone or even encourage QRPs. In the Discussion section, we reflect on this point by exploring the image of "pure science" in the business school.

"Of course, the peer-review process can also help to improve a paper and strengthen its contribution. But it is striking that scholars are willing to outsource their judgment to editors and reviewers insofar as the line between acceptable and unacceptable behavior is concerned."

DISCUSSION

The foregoing analysis shows that there are a number of QRPs in the business school, including playing with numbers, playing with models, and playing with hypotheses. Scholars account for these QRPs not only on an individual level (i.e., pressures and

incentives to publish), but also on an institutional level (i.e., insufficient methodological training, demands and expectations of journals). This serves to challenge the idea that QRPs can be avoided by enforcing stronger codes of conduct or making peer-review processes more stringent (e.g., Atwater et al., 2014; Chen, 2011; Harley, Faems, & Corbett, 2014; Schminke, 2009). In this section, we reflect on the "institutional and systemic structures" (Martinson, Anderson, & de Vries, 2005: 738) that play a role in cultivating an environment in which QRPs can take root. We also consider what can be done to minimize QRPs in the business school.

As we have seen, to label research practices as "questionable" does not mean that they are necessarily "bad;" rather, it means that certain practices fall into a gray zone between acceptable and unacceptable. For example, some respondents view altering leadership models as a "very serious problem," whereas others see it as "good exploratory research." Similarly, HARKing is seen by some as "a major violation of research protocol and research ethics," while others view it as "quite normal." Such practices are, therefore, contested terrain. This opens a discussion about the image of "pure science" in positivist management and organization studies.

Two ideals of science are prevalent in the business school. First, is the ideal of a unity of scientific method. In practice, this is associated with the hypothetico-deductive method, whereby theories are tested against "facts." This is precisely what makes replicability possible, which is regarded as the benchmark of any possible management science (Chen, 2011). Second, is the ideal of a fact-value distinction. This is expressed in the opening sentences of *The Leadership Quarterly* editorial published in wake of the recent retractions, which states that "leadership and leader performance are not a matter of opinion but rather a matter of fact" (Atwater et al., 2014: 1174). It is on this basis that leadership is viewed as a "scientific discipline" (*ibid.*).

In our view, the positivist image of science is demanding the impossible. After all, the hypothetico-deductive method was initially designed for the natural sciences where, under the right conditions, nature is much more likely to "speak back" to theories in a factlike manner than in the social world (Putnam, 2002). The desire to be a science "just like" the natural sciences means that journals are seeking to publish results that are seen to be objective and replicable.

The impossibility of living up to the positivist image of science is demonstrated by the way that scholars admit to playing with hypotheses. Ideally, the facts should be the test of a theory. One can do this by first formulating a theory and then, by means of rigorous application of methods, letting the facts speak for themselves—a process described by one respondent as "letting the chips fall where they do." As we have seen, however, scholars do not act in accordance with this ideal when they engage in HARKing. For some, HARKing is simply "bad science," because it disrupts the purity of the testing of theories. Others, however, view HARKing as a "reliable way of seeing the data," and therefore, a productive way of generating new insights, even if admitting to such practices in an academic article is liable to earn oneself an automatic rejection slip. This suggests that the ideal of science is more contested than academic journals would have us believe. The problem is that the traces of this contestation are effaced by the application of a QRP; that is, presenting a post hoc hypothesis as if it were an a priori hypothesis. As a result, scholars find themselves in a position where, to comply with the positivist image of science (i.e., hypothetico-deductive, value-free), they may end up negating this very ideal.

Such a paradox is also found in the other two forms of QRPs that we identified: playing with numbers and playing with models. Take, for example, the omission of outliers in statistical analyses. Our data suggest there is no unanimous agreement about how to deal with the identification and removal of outliers in leadership research. As one of our respondents puts it, there is always an undecidable question about "[w]hat's appropriate and what's not appropriate." Compounding this uncertainty is the suggestion that journals have certain expectations about how data sets "should" ideally look. Therefore, scholars may remove outliers to present a messy and complex social phenomenon in such a way that it begins to appear more objective and replicable than it really is. What results are studies in top journals that look "perfect," but which are "just not possible." Scholars thus engage in QRPs to produce studies that conform to an impossible image of science, which is seen as a prerequisite for publication in premier outlets. Likewise, while some scholars believe there are clear rules about how to measure and test models of leadership, others suggest that such rules are not agreed upon. The deciding

factor, often, is not whether it makes sense theoretically, but rather what statistically “produces ... a better result”—what is more likely to be more publishable in top-ranked journals. In both cases, we find that in the attempt to live up to the positivist image of science that appears in academic journals, researchers transgress this very ideal by engaging in QRPs.

It is tempting to point the finger of blame toward two commonly cited causes of QRPs: insufficient methodological training and pressures and incentives to publish. Indeed, many of our respondents turn to these explanations, as do commentators in management and organization studies (see e.g., Atwater et al., 2014; Chen, 2011; Harley, Faems, & Corbett, 2014; Schminke, 2009). However, focusing on the individual incentives and rewards that shape research practices neglects broader institutional forces in the business school.

A fuller account of research misconduct needs to acknowledge the role played by academic journals (along with senior editors and reviewers) in fostering QRPs. Our respondents tell us that leadership journals favor certain kinds of studies over others: studies that confirm hypotheses, studies that test new hypotheses rather than replicate existing studies, studies that show significant findings, and studies that adopt a certain method of measurement and analysis, to name but a few. This serves to restrict how research is designed, conducted, and reported, and in some cases, may incentivize authors to “package” their findings to suit the journal’s preferences. Some respondents reveal how these expectations are by no means subtle: Sometimes, reviewers and editors ask authors to “do things that are actually in a gray zone,” such as HARKing or being instructed to “change the method” in line with the editor’s own research agenda. The nature of this interaction means that authors willingly accede to such editorial requests to boost their chances of publication. What we are suggesting here is that journal editors and reviewers should not only be seen on the side of the angels, as gatekeepers of knowledge who tirelessly battle academic misconduct to protect the integrity of science, but precisely as part of the problem.

RECOMMENDATIONS

So, what should be done about QRPs in management and organization studies? Commentators

have proposed a range of specific measures to counteract academic misconduct, such as creating registry databases where researchers would submit their specific intentions (hypotheses, variables, estimated sample sizes, etc.) prior to conducting research; establishing a 2-step editorial review process to ensure that theory and methods are assessed separately from the research findings; encouraging the publication of null hypotheses and replications; improving systems of methodological training and ethical mentoring, particularly for early-career researchers; and supporting academic whistleblowers (Kepes & McDaniel, 2013; Kornfeld, 2012). These are no doubt useful guidelines for stimulating scholarly best practice or the “responsible conduct of research” (Steneck, 2006). However, the discussion around combating FFP predictably has been more vigorous than the discussion around reducing QRPs (if the distinction is made at all). This is understandable since FFP involves the unequivocal breach of scientific protocol, whereas QRPs are more difficult to categorically define and therefore prevent; moreover, whether we should prevent certain QRPs in all circumstances is itself open to question.

Nonetheless, we can take productive steps to reducing QRPs, or at least to prompt researchers to think twice about engaging in them. This is an important task because research has shown in other fields such as healthcare that “the greatest public harm ... stems from QRP, not FFP” (Steneck, 2006: 66), especially in terms of using resources wisely and making informed decisions about medical treatment and intervention. Part of the challenge, of course, is that QRPs remain hidden from view and are more widely tolerated than FFP, and cannot be explicitly punished (Fanelli, 2013a). Efforts to counteract QRPs will therefore require alternatives tactics to whistle blowing.

Many journals and professional bodies have a code of conduct that outlines appropriate and inappropriate behavior for its contributors or members. For example, the Academy of Management Code of Ethics (2006) provides a detailed list of enforceable ethical standards in research and publication, among other areas, covering both FFP and some QRPs (although it does not use this terminology). However, the problem with such guidelines is that they emphasize the behavior of individuals who may engage in QRPs, and neglect the role of journals in encouraging potentially inappropriate research practices.

One possible response would be to establish a "Transparency Index," which has been proposed by the founders of Retraction Watch (Marcus & Oransky, 2012). The Transparency Index would provide a numerical metric of the journal's transparency in different areas: how the journal conducts reviews (e.g., typical number of reviewers, acceptance rate); how the journal asks authors to disclose conflicts of interests; whether the journal requires authors to make their data set available; whether the journal uses plagiarism software; and how the journal deals with retractions (Marcus & Oransky, 2012). The Transparency Index could also measure whether journals provide an acknowledgment whenever an author changes their method or hypothesis at the editor or reviewer's suggestion. This would, at the very least, be a good start in shifting the emphasis from individuals to institutional structures. However, as Marcus and Oransky note, the Transparency Index—if it is implemented—is meant as an indicator, rather than a fool-proof measure, so that "[s]cientists' judgment will still be the most important factor" (Marcus & Oransky, 2012).

This brings us to the second possible response to QRPs, proposed by Fanelli (2013b): to redefine academic misconduct as "distorted reporting." For Fanelli, *distorted reporting* refers to "any omission or misrepresentation of the information necessary and sufficient to evaluate the validity and significance of research" (2013b: 149) in other words, as a disparity between what was done and what was reported. The benefit of this definition is that it captures not only FFP, but also QRPs. Typically, definitions of academic misconduct focus on the most egregious cases of fabrication, falsification and plagiarism, but tend to overlook less flagrant, more subtle forms of possible misconduct. This allows a whole series of practices to fall through the net and results in published work that is misleading in some way. Distorted reporting, by contrast, calls upon authors to explain and justify practices that fall into a potential gray area. For example, Fanelli (2013b: 149) gives the example of data dredging: "A scientist should be free to decide ... that 'fishing' for statistical significance is necessary. However, guidelines would require a list of every test used, allowing others to infer the risk of false positives." The same principle would also apply to HARKing: Authors should be permitted to hypothesize after the fact, but they will also need account for what they have done and how they

have done it, rather than simply concealing it from editors and reviewers. Likewise, if editors and reviewers ask an author to engage in HARKing, this too should be acknowledged in the final product.

Of course, some may feel this does not go far enough. Our reluctance to call for the outright prohibition of (some or all) QRPs, however, should not be mistaken for a lack of decisiveness on our part. Indeed, we feel that strict rules about QRPs may be counterproductive because they close down discussion about what is and what is not permissible in academic research, and more important, under what conditions. Crucially, redefining academic misconduct as distorted reporting encourages scholars to reflect on the standards of research in their field because they will be compelled to explain and justify research practices that may otherwise remain hidden from view. We feel it is therefore a positive step toward reducing QRPs, because it would start a conversation about precisely what information is "necessary and sufficient" in management and organization studies to evaluate the validity and significance of research. We assume that different norms of behavior and rules of conduct would apply to positivist research and interpretivist research. As such, shifting the focus from academic misconduct to distorted reporting would not be a smooth or easy process, but would necessarily involve a lengthy process of negotiation and "collective dialogue" (Fanelli, 2013b: 149) between editors, reviewers, and authors. This is precisely the kind of debate that we seek to open up here.

CONCLUSION

In this article we have taken seriously the call for more research that considers how QRPs are shaped by "institutional and systemic structures" (Martinson, Anderson, & de Vries, 2005: 738). By focusing on leadership scholars, we have discussed a number of QRPs in the business school. We see QRPs as "pinch points" (De Vries, Anderson, & Martinson, 2006: 50) that reveal a number of interesting dynamics. First, the very fact that there are research practices deemed questionable suggests that the positivist image of science in management and organization studies is more contested than many dare to admit. Second, by trying to live up to this image of science, some scholars engage in QRPs to produce research that (at least on the surface) resembles this ideal. This leads to our third

finding that academic journals—particularly editors and reviewers—may find themselves fostering QRPs, intentionally or otherwise.

To question the dominant image of science serves a larger purpose than addressing the problem of QRPs in the business school. Once we accept that the image of science demands the impossible, we may then ask, what is possible? This pertains not only to the ethics of a field of research (i.e., which research practices are acceptable and which are unacceptable), but more important to its very success. If we wish to develop as a social science, we must return to basic epistemological questions, such as “what kind of phenomena am I studying?” and “to what extent and by what means can I gain knowledge about this phenomenon?” Of course, to some extent this has already happened (e.g., Burrell & Morgan, 1979; Ciulla, 1995; Pfeffer, 1977). However, this kind of questioning should not be restricted to a few studies that deal with epistemological and ontological questions; it ought to be part of the research ethos of management studies as such, whether positivist, interpretivist, or critical. A debate about the positivist image of science in the business school will not erase QRPs. But it will lift the debate up to a higher level: what we can and cannot expect from different types of research. It would therefore be advisable for business school scholars to train more broadly in the philosophy, history, and sociology of the social sciences. What we need most are scholars who are willing to problematize dominant assumptions about how to study complex social phenomena like management, leadership, and organization, and it would be a shame if our journals continue to exclude those voices from the debate by insisting upon a particular image of science.

REFERENCES

- Academy of Management. 2006. *Code of ethics*. Retrieved from: http://aom.org/uploadedFiles/About_AOM/Governance/AOM_Code_of_Ethics.pdf.
- Alvesson, M., & Skoldberg, K. 2009. *Reflexive methodology: New vistas for qualitative research*. London: Sage.
- Anonymous 2015. The case of the hypothesis that never was: Uncovering the deceptive use of post hoc hypotheses. *Journal of Management Inquiry*, 24(2): 214–216.
- Atwater, L. E., Mumford, M. D., Schriesheim, C. A., & Yammarino, F. J. 2014. Retraction of leadership articles: Causes and prevention. *The Leadership Quarterly*, 25(5): 1174–1180.
- 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., 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.
- Bhattacharjee, Y. 2013, April 26. The mind of a con man. *The New York Times*. Retrieved from: <http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?pagewanted=all>.
- Budd, J. 2013. The Stapel case: An object lesson in research integrity and its lapses. *Synesis: A Journal of Science Technology, Ethics, and Policy*, 4: 47–53.
- Burrell, G., & Morgan, G. 1979. *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. London: Heinemann.
- Butler, N., Delaney, H., & Spoelstra, S. 2015. Problematizing ‘relevance’ in the business school: The case of leadership studies. *British Journal of Management*, 26(4): 731–744.
- Butler, N., & Spoelstra, S. 2012. Your excellency. *Organization*, 19(6): 891–903.
- Butler, N., & Spoelstra, S. 2014. The regime of excellence and the erosion of ethos in critical management studies. *British Journal of Management*, 25(3): 538–550.
- Butler, N., & Spoelstra, S. 2015. The dark arts of academia—and why journals must do more to tackle the problem. *The Conversation*, January 7. Retrieved from: <https://theconversation.com/the-dark-arts-of-academia-and-why-journals-must-do-more-to-tackle-the-problem-35796>.
- Carey, B. 2015, May 28. Study on attitudes towards same-sex marriage is retracted by a scientific journal. *The New York Times*. Retrieved from: <http://www.nytimes.com/2015/05/29/science/journal-science-retracts-study-on-gay-canvassers-and-same-sex-marriage.html>.
- Chen, X.-P. 2011. Author ethical dilemmas in the research publication process. *Management and Organization Review*, 7(3): 423–432.
- Ciulla, J. B. 1995. Leadership ethics: Mapping the territory. *Business Ethics Quarterly*, 5(1): 5–28.
- Chubin, D. E. 1985. Misconduct in research: An issue of science policy and practice. *Minerva*, 23(2): 175–202.
- Corbin, J., & Strauss, A. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory*, (3rd ed.). London: Sage.
- Davis, G. F. 2014. Editorial essay: Why do we still have journals? *Administrative Science Quarterly*, 59(2): 193–201.
- De Vries, R., Anderson, M. S., & Martinson, B. C. 2006. Normal misbehavior: Scientists talk about the ethics of research. *Journal of Empirical Research on Human Research Ethics: JERHRE*, 1(1): 43–50.
- 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. 2012. The good, the bad and the grey areas: Towards an international and interdisciplinary definition of scientific misconduct. In T. Mayer & T. N. Steneck (Eds.), *Promoting research integrity in a global environment*: 79–90. Singapore: World Scientific Publishing.

- Fanelli, D. 2013a. Why growing retractions are (mostly) a good sign. *PLoS Medicine*, 10(12): 1–6.
- Fanelli, D. 2013b. Redefine misconduct as distorted reporting. *Nature*, 494: 149.
- Fang, F. C., Steen, R. G., & Casadevall, A. 2013. Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42): 17028–17033.
- Fiedler, K., & Schwarz, N. 2015. Questionable research practices revisited. *Social Psychological & Personality Science*, 45–52.
- Gambino, L., & Devlin, H. 2015. Study of attitudes to same-sex marriage retracted over 'Fake Data'. *The Guardian*, May 20. Retrieved from: <http://www.theguardian.com/us-news/2015/may/20/same-sex-marriage-attitudes-study-retracted-fake-data>.
- Harley, B., Faems, D., & Corbett, A. 2014. A few bad apples or the tip of an iceberg? Academic misconduct in publishing. *Journal of Management Studies*, 51(8): 1361–1363.
- Harvie, D., Lightfoot, G., Lilley, S., & Weir, K. 2012. What are we to do with feral publishers? *Organization*, 19(6): 905–914.
- Harvie, D., Lightfoot, G., Lilley, S., & Weir, K. 2013. Publisher, be damned! From price gouging to the open road. *Prometheus*, 31(3): 229–239.
- 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.
- 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.
- Kepes, S., & McDaniel, M. A. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology*, 6(3): 252–268.
- Kerr, N. L. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3): 196–217.
- Kornfeld, D. S. 2012. Research misconduct: The search for a remedy. *Academic Medicine*, 87(8): 877–882.
- Lenz, H. 2014. Scientific ethics and publishing conduct. *Journal of Business Economics*, 84(9): 1167–1189.
- List, J. A., Bailey, C. D., Euzent, P. J., & Martin, T. L. 2001. Academic economics behaving badly? A survey on three areas of unethical behaviour. *Economic Inquiry*, 39(1): 162–170.
- Lynöe, N., Jacobsson, L., & Lundgren, E. 1999. Fraud, misconduct or normal science in medical research—An empirical study of demarcation. *Journal of Medical Ethics*, 25(6): 501–506.
- Macdonald, S. 2015. Emperor's new clothes: The reinvention of peer review as myth. *Journal of Management Inquiry*, 24(3): 264–279.
- Marcus, A., & Oransky, I. 2012. Bring on the Transparency Index. *Scientist (Philadelphia, PA)*, 1 (August). Retrieved from: <http://www.the-scientist.com/?articles.view/articleNo/32427/title/Bring-On-the-Transparency-Index>.
- Marcus, A., & Oransky, I. 2015, May 22. What's behind big science frauds? *The New York Times*. Retrieved from: <http://www.nytimes.com/2015/05/23/opinion/whats-behind-big-science-frauds.html>.
- Martin, B. R. 2013. Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy*, 42(5): 1005–1014.
- Martinson, B. C., Anderson, M. S., & de Vries, R. 2005. Scientists behaving badly. *Nature*, 435: 737–738.
- McNutt, M. 2015. Editorial expression of concern. *Science*, 348: 1100.
- Necker, S. 2014. Scientific misbehavior in economics. *Research Policy*, 43(10): 1747–1759.
- New York Times Editorial*. 2015, June 1. Scientists who cheat. *The New York Times*. Retrieved from: <http://www.nytimes.com/2015/06/01/opinion/scientists-who-cheat.html>.
- Panel on Scientific Responsibility and the Conduct of Research. 1992. *Responsible science: Ensuring the integrity of the research process*, Vol. I, Washington, DC: National Academy Press.
- Prasad, P. 2005. *Crafting qualitative research: Working in the postpositivist traditions*. NY: M.E. Sharpe.
- O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. 2017. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 43(2): 376–399.
- Office of Science and Technology Policy. 2000. Federal research misconduct policy. *Federal Register*, 65(235): 76260–76264.
- Pfeffer, J. 1977. The ambiguity of leadership. *Academy of Management Review*, 2(1): 104–112.
- Putnam, H. 2002. *The collapse of the fact/value dichotomy and other essays*. Cambridge, MA: Harvard University Press.
- Resnik, D. B. 2003. From Baltimore to Bell Labs: Reflections on two decades of debate about scientific misconduct. *Accountability in Research*, 10(2): 123–135.
- Resnik, D. B., Neal, T., Raymond, A., & Kissling, G. E. 2015. Research misconduct definitions adopted by U.S. research institutions. *Accountability in Research*, 22(1): 14–21.
- Retraction Watch. 2014a, February 7. Leadership journal to retract five papers from FIU scholar. *Retraction Watch*. Retrieved from: <http://retractionwatch.com/2014/02/07/leadership-journal-to-retract-five-papers-from-fiu-scholar>.
- Retraction Watch. 2014b, June 16. Ulrich Lichtenthaler retraction count rises to 16. *Retraction Watch*. Retrieved from: <http://retractionwatch.com/2014/06/16/ulrich-lichtenthaler-retraction-count-rises-to-16>.
- Retraction Watch. 2015, December 8. Diederik Stapel now has 58 retractions. *Retraction Watch*. Retrieved from: <http://retractionwatch.com/2015/12/08/diederik-stapel-now-has-58-retractions>.
- Schekman, R. 2013, December 9. How journals like *Nature*, *Cell* and *Science* are damaging science. *The Guardian*. Retrieved from: <http://www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science>.
- Schminke, M. 2009. Editor's comments: The better angels of our nature—Ethics and integrity in the publishing process. *Academy of Management Review*, 34(4): 586–591.
- Sovacool, B. K. 2008. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? *Journal of Bioethical Inquiry*, 5(4): 271–282.

- Steneck, N. 2006. Fostering integrity in research: Definition, current knowledge, and future directions. *Science and Engineering Ethics*, 12(1): 53–74.
- Stroebe, W., Postmes, T., & Spears, R. 2012. Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7(6): 670–688.
- Van Noorden, R. 2011. The trouble with retractions. *Nature*, 478: 26–28.
- Wenger, N. S., Korenman, S. G., Berk, R., & Berry, S. 1997. The ethics of scientific research: An analysis of focus groups of scientists and institutional representatives. *Journal of Investigative Medicine*, 45(6): 371–380.

Nick Butler is an assistant professor at Stockholm University Business School, Sweden. His research focuses on the sociology of work and the politics of knowledge production in the business school. He is a member of the editorial collective of *ephemera: theory & politics in organization*.

Helen Delaney is a senior lecturer in the Department of Management and International Business at the University of Auckland, New Zealand. Her research explores issues such as the role of work in creating and addressing social inequalities; the politics of science, relevance and commercialization in business research; and the intersections between discourse, power, and identity in work contexts. She is a member of the editorial collective of *ephemera: theory & politics in organization*.

Sverre Spoelstra is an associate professor at the Department of Business Administration, Lund University, Sweden. His research interests include organizational philosophy, the relation between work and play, and the use of psychometric instruments in management knowledge production. He is currently writing a book on the philosophy of leadership.