The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/2059-5794.htm

Received 31 March 2016
Revised 31 March 2016
Accepted 31 March 2016

REGULAR ISSUE PAPER

What happened to Popperian falsification? Publishing neutral and negative findings

Moving away from biased publication practices

Arjen van Witteloostuijn
Tilburg University, Tilburg, The Netherlands;
Antwerp Management School, University of Antwerp, Antwerp, Belgium and
Cardiff Business School, Cardiff University, Cardiff, UK

Abstract

Purpose – Current publication practices in the scholarly (International) Business and Management community are overwhelmingly anti-Popperian, which fundamentally frustrates the production of scientific progress. This is the result of at least five related biases: the verification, novelty, normal science, evidence, and market biases. As a result, no one is really interested in replicating anything. In this essay, the author extensively argues what he believes is wrong, why that is so, and what we might do about this. The paper aims to discuss these issues.

Design/methodology/approach – This is an essay, combining a literature review with polemic argumentation.

Findings – Only a tiny fraction of published studies involve a replication effort. Moreover, journal authors, editors, reviewers and readers are not interested in seeing nulls and negatives in print. This replication crisis implies that Popper's critical falsification principle is actually thrown into the scientific community's dustbin. Behind the façade of all these so-called new discoveries, false positives abound, as do questionable research practices meant to produce all this allegedly cutting-edge and groundbreaking significant findings. If this dismal state of affairs does not change for the good, (International) Business and Management research is ending up in a deadlock.

Research limitations/implications – A radical cultural change in the scientific community, including (International) Business and Management, is badly needed. It should be in the community's DNA to engage in the quest for the "truth" – nothing more, nothing less. Such a change must involve all stakeholders: scholars, editors, reviewers, and students, but also funding agencies, research institutes, university presidents, faculty deans, department chairs, journalists, policymakers, and publishers. In the words of Ioannidis (2012, p. 647): "Safeguarding scientific principles is not something to be done once and for all. It is a challenge that needs to be met successfully on a daily basis both by single scientists and the whole scientific establishment."

Practical implications – Publication practices have to change radically. For instance, editorial policies should dispose of their current overly dominant pro-novelty and pro-positives biases, and explicitly encourage the publication of replication studies, including failed and unsuccessful ones that report null and negative findings.

Originality/value – This is an explicit plea to change the way the scientific research community operates, offering a series of concrete recommendations what to do before it is too late.

Keywords Replication, Falsification, Publication bias, Research malpractices

Paper type Viewpoint

1. The human mind and science as a practice

Science is a community of human beings of the *homo sapiens* species: bipedals with the capacity to be self-reflexive[1]. This implies that science as a community is subject to all the same behavioral patterns that all human communities are, including a plethora of

Cross Cultural & Strategic Management
Vol. 23 No. 3, 2016
pp. 481-508
© Emerald Group Publishing Limited
2059-5794
DOI 10.1108/CCSM-03-2016-0084

biases at both the individual and collective level (Kahneman, 2011; Shleifer, 2012). Examples of well-known individual-level biases are hubris, confirmatory preference, and desire for novelty (or the reverse: fear of the new). This implies, for instance, that "When an experiment is not blinded, the chances are that the experimenters will see what they 'should' see" (*The Economist*, 2013). Together, these biases lead to Type I and Type II errors in judging research, both our own and that of others[2]. As a result, without correcting mechanisms, published research will be heavily biased in favor of evidence that is in line with the theory.

Science's first line of defense is the micro-level reviewing process. Regrettably, the reviewing process, double-blinded or not, is anything but flawless, but rather full of biases itself (for a critical discussion regarding Business and Management, see e.g., Bedeian, 2003; Starbuck, 2003; Tsui and Hollenbeck, 2009). This is not surprising, as the reviewing process is carried out by exemplars of the very same *homo sapiens* species that cannot escape from all these biases referred to above (plus quite a few others). Particularly in Medicine, ample evidence abounds that current reviewing practices fail to deliver the effective filtering mechanism they are claimed to provide (see e.g. Jefferson *et al.*, 2002). Take the revealing study of Callaham and McCulloch (2011). On the basis of a 14-year sample of 14,808 reviews by 1,499 reviewers rated by 84 editors, they conclude that the quality scores deteriorated steadily over time, with the rate of deterioration being positively correlated with reviewers' experience. This is mirrored in the well-established finding that reviewers, on average, fail to detect fatal errors in manuscripts, which reinforces the publication of false positives (Callaham and Tercier, 2007; Schroter *et al.*, 2008).

Hence, giving these unavoidable biases associated with the working of the human brain, the scientific community should adhere, as a collective, to a set of macro-level correcting principles as a second line of defense. Probably the most famous among these is Popper's falsifiability principle. Key to Karl Popper's (1959) incredibly influential philosophy of science is his argument that scientific progress evolves on the back of the falsification principle. We, as researchers, should try, time and again, to prove that we are wrong. If we find the evidence that indeed our theory is incorrect, we can further work on developing new theory that does fit with the data. Hence, we should teach the younger generation of researchers that instead of being overly discouraged, they should be happy if they cannot confirm their hypotheses[3]. This quest for falsification is critical because, in the words of Ioannidis (2012, p. 646), "Efficient and unbiased replication mechanisms are essential for maintaining high levels of scientific credibility." The falsification principle requires a tradition of replication studies in combination with the publication of non-significant and counter-results, or so-called nulls and negatives, backed by systematic meta-analyses[4].

2. The medical trial and social psychology examples

The failure of both lines of defense would frustrate scientific progress, undermining the credibility of the scientific community. This is a recurring issue in Medicine. In Medicine, the bias in publication practices regarding clinical trials is a deeply rooted problem, triggering a permanent debate as to the credibility of the published evidence, and about what can be done to solve the issue. Many argue that the conflict of interest due to the heavy involvement of the multi-billion pharmaceutical industry greatly reinforces the publication bias that already "naturally" emerges as a result of the toxic mixture of academic incentives and human biases (see e.g. Lexchin *et al.*, 2003; Melander *et al.*, 2003): (false) positives are published, while nulls and negatives are not.

The recent AllTrials initiative (www.alltrials.net) is an example of an attempt to counter this credibility-undermining tendency: "The AllTrials campaign calls for all past and present clinical trials to be registered and their results reported." The AllTrials international initiative's petition had been signed by 85,424 people and 597 organizations on July 29, 2015.

In Psychology, a discipline closer to Business and Management, a number of headlines-hitting scandals in the early 2010s triggered Economics Nobel Prize winner Daniel Kahneman to publish an open letter asking for a swift and forceful response, as reported in *Nature* (www.nature.com): "The storm of doubts is fed by several sources, including the recent exposure of fraudulent researchers, general concerns with replicability that affect many disciplines, multiple reported failures to replicate salient results in the priming literature, and growing belief in the existence of a pervasive file-drawer problem that undermines two methodological pillars of your field [social priming: AvW]: the preference for conceptual over literal replication and the use of meta-analysis. […] For all these reasons, right or wrong, your field is now the poster child for doubts about the integrity of psychological research" (Kahneman, 2011, September 26).

In response to this crisis, a number of promising initiatives have been launched, of which three are particularly worth emphasizing here, by way of illustration. First, the "Reproducibility Project: Psychology" is a crowdsourced empirical endeavor of 270 scholars from around the world producing 100 replication studies (https://osf.io/ezcuj/wiki/home/). In September 2015, the results were published in *Science*, revealing that of the original 97 percent of positives, only 36 percent survived the replication endeavor (Open Science Collaboration, 2015). Second, the Many Labs Replication Project was launched to examine whether or not 13 well-known psychological studies could be replicated (https://osf.io/wx7ck/). By the end of 2013, this group had successfully replicated the results of 10 out of 13 prior experiments (Yong, 2013)[5]. Third, in 2015, the *Comprehensive Results in Social Psychology* journal was established by the European Association of Social Psychology to not only publish replications and extensions (next to original research), but also to work with a two-step reviewing process, "using the registered report format where a plan for research is submitted for initial review. […] If the plan for research is accepted […] authors are guaranteed publication of the manuscript irrespective of the outcome of data analysis" (www.tandfonline.com/loi/rrsp20#.VbtCRDcw-70; see also the launch editorial by Jonas and Cesario, 2015).

3. The verification bias[6]

The debate in Medicine and Psychology largely relates to experimental studies conducted in a laboratory (or lab, by way of shortcut) or guided by a strict trial protocol. And related to this, the countermeasures involve the registration and/or replication of such lab and/or trial work. What makes our field, broadly referred to as (International) Business and Management[7], very different from Medicine and Psychology is that by far the majority of the empirical studies make use of noisy data collected outside of a lab context and without a strict trial protocol[8]. With the exception of organizational behavior and marketing, the bulk of empirical research in all Business and Management's sub-disciplines, from entrepreneurship and strategy to management accounting and organization theory, involves non-random field data, with lab and field experiments being the exception rather than the rule (van Witteloostuijn, 2015). Inevitably, this implies that the issues in the context of the Business and

Management field can be different, as may be the workable countermeasures, *vis-à-vis* those in Medicine and Psychology. However, at their root, the issues and measures are not fundamentally different, I would argue, because the same biases have affected the Business and Management scholarly community.

One bias that stands out in Business and Management (but not only there – quite the contrary; see above, and more on that below), is very anti-Popperian in nature. As Business and Management's scholarly community, falsifiability is not the principle that we adhere to all the time. There appears to be a disconnect between what we practice from what we preach. In reality, we are obsessively focussed on the verification principle: that is, on trying to prove that we are right by generating positives. In so doing, we claim to be Popperian while consistently violating the very principle that is the cornerstone of Popper's view as to what drives scientific progress. Even worse, this may inadvertently incentivize malpractices that run counter to what science should be all about[9]. How often have we, as serious scholars, turned our theory upside down such that we could report that – hurray! – our analyses confirm many, if not all, of our hypotheses? This malpractice of so-called HARKing (= hypothesizing after the results are known; Kerr, 1998) may be inadvertently stimulated by the behavior of many editors and reviewers. If we submit a paper in which we, basically, admit that our original theory cannot be confirmed by the data, many editors/reviewers may either decide/advice to reject the paper, or suggest identifying alternative theories such that the findings are more in line with the predictions[10].

Faced with a "publish or perish" culture[11], it is hard to resist the temptation to adapt to the anti-Popperian rules of the modern and highly competitive publication game. We all need publications to find an (entry) job, to obtain tenure, to be promoted, to receive a higher salary, and/or to gain status. Hence, as most journals' (often tacit) publication criteria are geared toward verification, our journals are filled with studies that happily report evidence that is in line with the suggested theory. How often do we see a paper in which all hypotheses are rejected, with a discussion section in which new theory is developed that fits with the original theory's non-evidence? How many papers have a back-end in which the study's real theoretical action is revealed in the form of a series of *post hoc* analyses that are meant to explore new explanations for the many non- or contra-findings – or nulls or negatives – that were produced on the basis of the original theory? Asking these questions implies answering them (Bettis, 2012). Our field, and not our field alone, suffers from a very serious verification bias.

4. The novelty bias

Aside from a verification bias, another common publication practice that frustrates scientific scrutiny is the endemic demand for novelty (cf. Mittelstaedt and Zorn, 1984) and new theory (Van Maanen, 1989). The mission statement of the Business and Management's premier empirical outlet, the *Academy of Management Journal*, does not leave any room for doubt: "Authors should strive to produce original, insightful, interesting, important, and theoretically bold research. Demonstrations of a significant 'value-added' contribution to the field's understanding of an issue or topic is crucial to acceptance for publication." This is not different for the other top outlets for empirical Business and Management work: "Look to ASQ for new work from young scholars with fresh views, opening new areas of inquiry, and from more seasoned scholars deepening earlier work and stalking out new terrain" (*Administrative Science Quarterly*); "*JOM* encourages new ideas or new perspectives on existing research" (*Journal of Management*); "*JMS* publishes innovative empirical and conceptual articles which advance knowledge of management and organization broadly defined" (*Journal of Management Studies*); and "the journal publishes groundbreaking research about organizations" (*Organization Science*). And so on and so forth. All journals want the same!

Apparently, we all love "bold theory," "breaking new ground," and "innovative research." However, such an obsessive focus on novelty should not be the sole engine of scientific progress. As in any discipline, real "groundbreaking research" is very rare, as are successful entries into "new terrain." Much scholarly work involves transpiration, and not so much inspiration. More importantly, this incremental[12] work is highly needed (cf. Helfat, 2007; Mezias and Regnier, 2007). In the world-class Barcelona football team, Lionel Messi's brilliance can only shine due to the key support of many complementary (and much less talented) teammates passing him the ball, as could Johan Cruijff's or Diego Maradonna's in the past. If Barcelona could field 11 Messi clones (which they cannot, given the scarcity argument), they would start losing one game after the other, as cooperation among complementary players is no longer possible in a team of 11 look-alikes, however brilliant. Science is not so different from football in this respect. Specifically, incremental, or rather: groundlaying, contributions are essential for at least three reasons.

The first reason is that we rarely[13] know *a priori* what research will turn out to be cutting-edge and path-breaking; if we could, academic life would be far easier, as we could cherry-pick what will work out brilliantly up-front, and would be way more boring, as the outcomes of our work will be predictable. Scientific (sub-)disciplines are eco-systems that benefit from effective selection processes. By producing much *ex ante* variety, *ex post* selection can do its beneficial work. After the fact, the scientific community can find out which "novelties" are really novel, and what "new ground" really turns out to be fertile. The second reason is that groundlaying contributions produce the building blocks for cutting-edge research that enters this new ground. The next kid on the block is not an alien, but rather combines features of many other ordinary kids into an exciting novel appearance. Oliver Williamson, for example, could not have developed his impressive transaction cost economics program without the groundlaying work executed by an army of hundreds of transaction cost scholars, across a wide variety of (sub-)disciplines. Similarly, Michael Hannan and John Freeman's organizational ecology puzzle is full of groundlaying pieces produced by dozens of loyal followers, to provide the fuel to keep the organizational ecology engine up and running.

The third reason involves an essential component of science's machinery: the key role of replication studies, both failed (conflicting) and successful (confirming) ones. The dominance of the so-called "file-drawer problem" (Rosenthal, 1979) would kill scientific progress, as "Replication […] is meaningless unless failed replications are published as enthusiastically as are successful replications" (Ferguson and Heene, 2012, p. 555). A scientific community that only generates novelty, cannot produce a solid knowledge base. Only by replicating prior work, can we develop external and internal validity; only then, the boundary conditions of earlier novelty can be defined, step by step (Bettis, 2012). This is even more important (and difficult) in many of the social sciences, including Business and Management, with much field data that are, in one way or another, unique. Designing a new radical architectural design on paper (or rather, on a computer screen) is one thing; building the structure with real concrete and wood is quite another matter. The precise material and the detailed design may well depend upon environmental circumstances, such as the nature of the soil conditions and the likelihood of flooding or earthquakes. Are transaction costs of a different nature in developing countries? Do ecological processes work out similarly in Cuba and China?

However, as a journal, recognizing the need for replication studies and successfully implementing a policy to really publish this type of work is easier said than done. An illuminating case in point is the *Strategic Management Journal* [14]. On the one hand, the *Strategic Management Journal* is very explicit about the danger of false positives and the need for replication work, as is clear from the study of Hubbard *et al.* (1998) and the essay by Bettis (2012)[15]. On the other hand, an electronic search for replication studies in the journal's e-archive (covering the 1980-now period) only gives two hits, after removing articles involving replication as a substantive issue (e.g. of knowledge or routines): Barker and Mone (1994), and Mayer and Whittington (2003)[16]. In other prominent Business and Management journals, the score is not much better. Take the *Administrative Science Quarterly* and *Organization Science*. After an electronic search in their e-archives (running from 1999 to 1990 until today, respectively, for the *Administrative Science Quarterly* and *Organization Science*), not a single hit emerges for "replication" as a keyword or in the title (and, for the *Administrative Science Quarterly*, in the abstract). This zero-hit result for the *Administrative Science Quarterly* is particularly striking, given that their website explicitly states that "Contributions can include [...] the disconfirmation of existing theory"[17].

In light of the above, the time is right for launching a *Journal of Business and Management Replication Studies*[18]. Without replication studies, successful and failed ones [19], Business and Management will end up in a deadlock. I suggest at least three types of replication studies (cf. Tsang and Kwan, 1999)[20]. The first type is that other scholars replicate a published study's analyses with this study's data[21]. This implies that the original raw data must be made available, as well as a manual carefully describing all data-handling procedures, including statistical routines[22]. A new standard practice could be that this material has to be uploaded to the journal's or an archive consortium's webpage, as part of the set of conditions associated with acceptance for publication. The second type involves replication studies with new data. This requires perfect mimicking of the original study's research design, implying that the latter's description has to be crystal-clear. Replicating with new data, from another sample (say, different individuals, teams, organizations, countries, and/or industries), is needed to explore the robustness of the earlier findings, which is instrumental in defining the original theory's boundary conditions. The third type bridges the world of replication with that of novelty: replications with extensions (cf. Hubbard and Vetter, 1996), in which a baseline model explicitly replicates earlier work before adding novelty in additional specifications[23].

5. The normal science bias

Thomas Kuhn (1962) famously inspired the introduction of the term paradigm shift to explain that normal science is a very different beast from its non-normal kin. Normal science is operating within the strict boundaries of a given and dominant paradigm. In normal science, findings that seriously go against the prevailing paradigm are not welcomed as a step toward further progress, but rather are put aside as mistakes of the researcher. A researcher either sticks to the rules of the prevailing paradigm and behaves nicely in line with the many explicit rules and implicit codes that regulate the paradigmatic field, or s/he is shunned and ignored by this normal science community. Again, a scientific community is not that different from any other community of the *homo sapiens* species. At a meta level, Business and Management seems to behave as if evolved into, in Kuhnian parlance, a normal science community. Of course, Business and Management hosts a plethora of theories, designs, methods, schools of thought, and research traditions. However, by and large, there is a solid world of conformity behind this impressive variety, as reflected in the verification and novelty biases.

Normal science is ruled by a variety of taken-for-granted norms and, often implicit, codes of conduct, many of which actually are urban myths. For example, in an earlier editorial (van Witteloostuijn, 2015), I already hinted at a few of these urban myths that circulate in the International Business (and Management) community, such as a dislike for mathematical modeling and student samples because these are argued to be associated with low external validity. But there are many more powerful do's and do not's. Another prominent double-do in Business and Management is that a study is only taken seriously if the reference list continues over many pages and if the theoretical contribution is non-incremental – quite a challenge indeed. The first leg of this double-do implies that whatever novelty is being produced in the second leg, its ingredients should be deeply grounded in the extant literature: only then, the paper can walk nicely on two legs. However, real novelty of the path-breaking kind often mixes new ingredients not heavily rooted in prior work.

Take Economics Nobel Prize winner Robert Coase's cutting-edge masterpiece: "The Nature of the Firm," 20 pages published in 1937 in *Economica*. Without this little essay, Williamsonian transaction cost theory, which is incredibly influential in Business and Management, would probably not have emerged – certainly not in its current form and shape. Strikingly, Coase's (1937) essay has no reference list[24]. All references are in the main text and in footnotes, all involving a real engagement with the referenced piece of work, rather than the lip-serving referencing that is normal practice in Business and Management (a typical example being a reference to DiMaggio and Powell's (1983) isomorphism in papers engaging with institutional theory; cf. Sutton and Staw, 1995). Business and Management's extensive referencing practices imply that a really novel argument is a hard sell. A little thought experiment (without any reference) should make this clear. What would be the response of reviewers if a submitted paper includes a lead-up to a new set of hypotheses with only a few references, or even without a single reference altogether? Two other examples of classic normal science practices, among many, are the requirements to spell out the "underlying theoretical mechanisms" and to present an "overarching theoretical framework."

Of course, all this is not to say that normal science practices are inherently dysfunctional – they are not. However, if overly dominant, such practices impede out-of-the-box thinking – not only regarding how to conduct empirical studies, but also how to develop theory (cf. Delbridge and Fiss, 2013). This implies that a second new journal is needed: the *Journal of Heterodox Perspectives in Business and Management*[25]. In an outlet like that, freestyle essays can be published without the requirement to extensively reference and pay due respect to the state of the art, and with the leeway to introduce non-conventional arguments and to apply non-orthodox methods. This requires a very different reviewing style, away from the multi-page list of – often very detailed – normal science comments. A short review of, say, half a page will do, answering only two key questions: are the ideas and/or findings presented thought-provoking?; and is the argumentation and/or analysis well done?[26].

6. Is Business and Management a normal science?

The above series of arguments and claims is not without intrinsic tension – possibly paradoxes, or perhaps even contradictions. Is Business and Management a normal science?

If so, how does the need for extensive referencing sit with the obsessive search for cutting-edge novelty? And will a *Journal of Heterodox Perspectives in Business and Management* not attract much junk science? In a normal science community, the dominant paradigm offers a guiding template for positioning studies and interpreting findings without the need to engage in extensive referencing. After all, the normal science paradigm comes with a set of common knowledge assumptions and theories that make extensive referencing redundant. For instance, in Economics, references to back up the assumption than an economic agent is maximizing a utility function is futile, and would even discredit the scholar doing so as an unknowledgeable outsider. Given this observation, one could argue that, by and large, Business and Management as a discipline is on its way to develop into a normal science, but is not there yet. Or, alternatively, Business and Management, being a set of disciplines rather than a mono-discipline, can never be a unitary normal science community by its very multi-disciplinary nature (cf. Pfeffer, 1993; Corley and Gioia, 2011)[27].

Whatever the reason for the current state of affairs, Business and Management hosts a peculiar mixture of normal and non-normal science practices. As a result, the major journals' messages are full of ambiguity. On the one hand, the "breaking new ground" and "cutting-edge" requirement signals a quest for work that breaks with the current paradigm(s). On the other hand, the extensive, if not excessive, referencing should tie this novelty to prevailing paradigm(s). The outcome is what may be referred to as a façade of "incremental cutting-edge" research, elaborating and fine-tuning existing theory, which comes with an emphasis on "fetishistic theory" (and methods, for that matter) at the expense of novel attention to important empirical phenomena (cf. Birkinshaw *et al.*, 2014; Colquitt and Zapata-Phelan, 2007; Davis, 2015; Hambrick, 2007; Helfat, 2007; Tsui, 2013). This obsession with novelty and theory in combination with extensive referencing comes with a publication straightjacket: many papers are look-alikes, many claiming to offer a major contribution to theory (which comes with much over-claiming and window-dressing arguments), many offering an (of course, supportive) empirical test of their own theory, many discussing the implications for extant theory, many listing managerial implications, etc. But actually, by far the majority of those papers in the *Academy of Management Journal*, *Administrative Science Quarterly*, *Strategic Management Journal* and all these other top journals looking for groundbreaking novelty tend to report groundlaying contributions deeply and squarely positioned in the large stock of related prior work (cf. Miller, 2007)[28].

A final interim remark relates to the function of officially publishing and referring to working papers. In a mature normal science, this is standard practice; in Business and Management, this is the exception rather than the rule. That is, in Business and Management, the norm is the hide-until-published approach: "do not publish until you really publish"[29]. This has at least two serious downsides. First, unpublished nulls and negatives are far more difficult to find, which makes conducting proper meta-analyses a much harder endeavor (see below). Second, authors, editors and reviewers can tweak the original study behind closed doors until a final version finds its published shape, full with all these nice positives. Here, Business and Management can learn from a sister discipline like Economics, with prominent and highly cited working paper series such as EconPapers, NBER, and SSRN.

7. The evidence bias

So, given the above, what is the nature of the evidence published in all these Business and Management journals, one quarter after the other? If three such fatal biases appear to "plague" publication practices, what then can we learn from the published evidence? The heart of the problem here is: there is no way we can be sure of any answer to this vital question. If only positives are published without any serious attempt at replication, the only thing we can be sure of is that false positives will abound. The title of John Ioannidis essay in *PLoS Medicine* of 2005 is revealing: "Why most published research findings are false" – a statement, not a question. His conclusion, as summarized in the abstract, is particularly worrisome for a field like Business and Management: "a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; when there is greater flexibility in designs, definitions, outcomes, and analytical models; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance[30]. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true" (Ioannidis, 2012, p. 0696; cf. *The Economist*, 2013).

Two additional, and closely connected, downsides of modern scientific practices provide further reason for pessimism. First, given the omnipresent chase for novel positives, we cannot expect that the nature of unpublished work will be that much different from what is being published. In all likelihood, unpublished work suffers from the very same pair of verification and novelty biases. Second, given the scientific community's codes and incentives, scholars either do not publish nulls or negatives at all, or engage in dubious research practices to generate the novel positives they are looking for – e.g., by " 'cleaning' data and rerunning analyses until expected results are achieved or by running simplistic analyses that favor one's own hypotheses" (Ferguson and Heene, 2012, p. 556). As a result, to paraphrase Ferguson and Heene, there is an 800-pound gorilla in Business and Management's living room. To deal with this challenge, the least we can do to combat this gorilla is to stimulate meta-analyses.

Hence, Business and Management needs yet another outlet: a *Journal of Business and Management Meta-Analyses*. Meta-analyses are well-established in fields such as Medicine and Psychology, but less so in Business and Management[31]. It is not that meta-analyses are absent altogether in Business and Management, but rather that such analyses are quite rare. A search in the e-archives of the *Academy of Management Journal* (1958-now), *Administrative Science Quarterly* (1999-now), and *Journal of Management* (1975-now) generated 14, 0, and 29 hits, respectively – not that many over so many decades. One reason for this follows from the above diagnosis: given the bias toward novel positives, much published Business and Management work tends to be unique in one way or another, making it very hard to collect a large enough number of studies that is sufficiently similar to carry out a quantitative meta-analysis. For that to be possible, a research tradition must share a paradigmatic core, such as is the case with organizational ecology (see e.g. Bogaert *et al.*, 2016). But for those topics within Business and Management that do share such a common core, carrying out meta-analyses would be highly valuable[32]. An example is reflected in the many foreign entry and establishment mode studies inspired by transaction cost theory (see e.g. Dikova and van Witteloostuijn, 2007). Another case in point is research on cultural diversity.

Regrettably, the above is easier said than done. For one, the number of cumulative stocks of sufficiently similar studies might well turn out to be rather limited. Moreover, finding non-biased unpublished work to correct for the pro-positives publication bias is likely to be problematic. But this does not imply that the Business and Management community should not start fighting this uphill battle. Initially, three weapons can be employed in the short run to manoeuver our field in a winning position on this battleground. First, to the extent feasible, further meta-analyses can be carried out, applying modern publication bias-correcting techniques (see below). Second, reporting practices in Business and Management should be improved such that studies include the data and information needed to construct meta-analytic data sets. Third, meta-analytic reviews without formal meta-analyses, as published quarterly in an outlet like *International Journal of Management Reviews*, may provide a steppingstone for the systematic accumulation of Business and Management knowledge (but see Stanley, 2001, for a critical discussion).

8. The market bias

That Business and Management, as in any other (sub-)discipline, is susceptible to this set of four, related, biases is fully understandable, given the nature of the "market" for academic journals, and the incentives that dominate the scholarly labor market. Regarding the former, the Holy Grail is impact (cf. Tsui, 2016). Journals want impact, as much as their editors and publishers do. Impact generates high-quality reputation, this reputation attracts high-quality submissions, and high-quality outlets can command high royalties. Quite strikingly, the global scientific community is deeply influenced by the commercial practices of a US stock-listed enterprise: Thomson Reuters. The first lines of the opening page of their (in)famous Web of Science (WoS) includes seven "this is ours, and no one else's!" sign marks (TM, ® and SM). Publishers of scientific journals, with Elsevier in a leading position, are addicted to WoS's Impact Factor race, notwithstanding attempts to come up with alternatives (such as Elsevier's Scopus and Google's Scholar).

In a commercial world like that, a different logic operates from what should drive the scientific community[33]. Editorial boards of journals, including those running Business and Management outlets, are seeking to increase their WoS Impact Factor, with a close eye on the performance of their competitors. Hence, they seek to publish papers that they expect will be much cited. For sure, replication studies fail to pass this hurdle. Instead, "cutting-edge" and "ground breaking" studies are in high demand. No wonder that all these journals' editorial policies are, in their core, look-alikes. Annually or semi-annually, editorial boards discuss strategies that can boost their journals' Impact Factor. How many special issues should be launched? Who should be invited to sit on the board? What type of special sections may be created? Should the use of social media be professionalized? etc.

The tension between these competing commercial *vis-à-vis* non-commercial logics is clearly reflected in the ongoing and heated debate about open access publishing. The Public Library of Science (PLoS) is the figurehead of the open access movement. The opening sentence of PLoS's website announces that "PLoS is a nonprofit publisher and advocate of open access research." The ideal is to escape from the dominant commercial paradigm of the likes of Elsevier and Thomson Reuters, and to provide a platform for a free world of nonprofit-motivated scientific exchange. However, PLoS is fully engaged, voluntary or involuntary, in the WoS Impact Factor game, too. For instance, *PLoS Biology* is number 2 of 85 in Biology with an Impact Factor of 9.3, *PLoS Medicine* is number 7 of 153 in general and internal Medicine with an Impact Factor of 14.4, and *PLoS ONE* is number 8 of 56 in multi-disciplinary sciences with an Impact Factor of 3.2 (WoS Journal Citation Reports 2014 Edition)[34] – all quite impressive indeed. Hence, the open access movement does not (yet) really solve the types of problems that are highlighted in this essay's central argumentation[35].

A very tricky and highly persevering issue is that modern scientific incentives are perverse. This implies that the chase for novel positives, and its concomitant dislike for nulls and negatives, is deeply institutionalized in the workings of the academic community (cf. Bedeian *et al.*, 2010). Making a career by performing replication studies or by reporting nulls and negatives is highly problematic, if not plainly impossible. So, this implies another market bias – that of the academic labor market. As Pashler and Wagenmakers (2012, p. 529) rightly observe, "the replicability problems will not be so easily overcome, as they reflect deep-seated human biases and well-entrenched incentives that shape the behavior of individuals and institutions." Indeed, the problem is grounded in the way any new generation of scholars is socialized: "Young investigators are taught early on that the only thing that matters is making new discoveries and finding statistically significant results at all cost" (Ioannidis, 2012, p. 647). So, the biases are deeply ingrained in institutionalized cultures, incentives and practices[36]. If there is one stylized fact in organization theory, then this is the one: changing such an institutionalized configuration is very hard indeed[37].

One final issue worth discussing is, given intrinsic human biases, the low-status nature of replication work[38]. In the words of Kane (1984, p. 3), "uninventively verifying someone else's research is not a completely respectable use of one's time. Choosing such a task is widely regarded as prima facie evidence of intellectual mediocrity." This implies that extra efforts are needed to get this off the ground, assuming that this aspect of the market bias cannot be changed so easily – if at all. One option is to involve novices. For instance, carrying out replication studies can be made part of training modules in PhD programs[39]. This makes perfect sense, as learning-by-replicating is an effective steppingstone for developing the many tacit skills involved in academic work. Another route is to launch, on a regular basis, tailor-made orchestrated international replication projects, similar to the Reproducibility Project: Psychology and Many Labs Replication Project referred to above. In both cases, further incentives emerge if the replication studies are of the third kind, including extensions, which are easier to publish. Outside the lab, such extensions are very likely to be part of replication work anyway, as one-to-one mimicking of social sciences' field work is close to a mission impossible.

9. *Déjá vu*

Clearly, the above diagnosis is anything but new – quite the contrary. In Economics, for instance, attention for publication bias is limited, with replication studies being as unpopular as in any other field in the social sciences. However, that replication studies are a necessary condition for scientific health is widely recognized in Economics (e.g. Blaug, 1992; Kane, 1984), as is the persistent problem of the pro-positives publication bias (e.g. Bradford De Long and Lang, 1992; Leamer, 1978). A key message from the Economics' publication bias literature is that meta-regression with correction for publication bias is essential, applying techniques such as funnel-asymmetry testing, meta-significance testing, and precision-effect testing (e.g. Stanley, 2001, 2008). Without such a correction, and given the lack of widespread availability of unpublished studies with nulls and negatives, meta-regression analysis will simply mimic the publication bias in the original material: "if uncorrected, meta-analysis is itself susceptible to the distortion of publication selection" (Stanley, 2008, p. 105). For instance, Doucouliagos and Stanley (2009) re-do Card and Krueger's (1995) meta-analysis of the minimum wage – employment relationship after adding a publication bias correction, and find that "Once this publication selection is corrected, little or no evidence of a negative association between minimum wages and employment remains" (Doucouliagos and Stanley, 2009, p. 406).

Psychology was probably the first discipline that started to take the publication bias issue seriously (Sterling, 1959). Since then, matters moved from bad to worse, as the discipline was hit by a series of scandals in the 2010s, with the Diederik Stapel affair as a dismal low. In fact, in a 2012 article in *Psychological Science*, John, Loewenstein and Prelec conclude that many psychologists admitted to engaging in at least a few questionable research practices in an attempt to produce the positives required to have their work accepted for publication. These questionable research practices run from relatively innocent (e.g. data-mining and outlier removal) to outright fraud (e.g. data fabrication and voodoo correlations). And even if the research practices are non-questionable, the way of reporting often is not (Wigboldus and Dotsch, 2015)[40]. These are serious violations as they corrupt the credibility of accumulated evidence, frustrate conducting valid meta-analyses, and undermine the community's integrity (cf. the painfully funny two-pager of Neuroskeptic, 2012 – *mea culpa*).

In Medicine, Begg and Berlin (1988, p. 420) already observed the following almost three decades ago: "However, among those who are active in clinical research there is […] in extreme cases […] an attitude of almost limitless cynicism and incredulity regarding published clinical reports, especially reports of new treatments, and especially studies involving non-randomized design. Among such participants there is a belief that the presence of a significant result is merely a necessary attribute for persuading journal editors that a paper is worth publishing, rather than a realistic probabilistic summary of the inference regarding the hypothesis under study." Decades later, this issue is still on the table, notwithstanding new meta-analytic tools available for the systematic evaluation of cumulative research findings after correction for publication bias. Examples of such tools are funnel plot techniques and regression-based adjustment methods, similar to those promoted in Economics (e.g. Sterne *et al.*, 2001; Moreno *et al.*, 2009, 2011; Million and Raoult, 2012).

Still, the prominent advocate of fighting the battle against the publication bias in Medicine and beyond, John Ioannidis (e.g. 2005 and 2012), has relentlessly warned, and still does so, the scholarly community that the current state of affairs is depressing. Journals are not interested in publishing nulls and negatives, let alone in failed replication studies, neither are funders, institutes and researchers. The danger of this should not be underestimated, as is evident in Ioannidis' (2012, pp. 646-647) rhetorical question: "could it be that the advent of research fields in which the prime motive and strongest focus is making new discoveries and chasing statistical significance at all cost has eroded the credibility of science and credibility is decreasing over time?" This prime motive is bred in an environment full of perverse incentives, implying that "No one is interested in replicating anything" (Ioannidis, 2012, p. 647). In many sciences, independent replication studies represent only one or two percent of published articles (e.g. Evanschitzky *et al.*, 2007; Makel *et al.*, 2012). The majority of the publications involve "unchallenged fallacies" (Ioannidis, 2012). The result is that all the lip service paid to Popper's principles is a façade. Behind this façade, the machinery of scientific progress is seriously eroded.

Regrettably, this depressing observation holds true for Business and Management, broadly defined, too. On the basis of a careful analysis of 4,270 empirical studies published in 472 issues of 18 top journals in Accounting, Economics, Finance, Management, and Marketing over the 22-year time period running from 1970 to 1991, Hubbard and Vetter (1996, p. 153) report that replications (with or without extensions) "typically constitute less than 10 percent of published empirical work in the accounting, economics, and finance areas, and 5 percent or less in the management and marketing fields." Given these meager percentages, they came to the dismal conclusion that: "At present, many empirical findings in the business literature are isolated and fragile" (Hubbard and Vetter, 1996, p. 161). Regrettably, this "At present" disclaimer does not offer an escape. In 2007, Hubbard, with three other co-authors, updated part of their counting exercise, with the frightening result that the replication rate had fallen to a new low of 1.2 percent in leading Marketing journals (Evanschitzky *et al.*, 2007).

Moreover, as Bettis (2012, p. 110; see also e.g. Sutton and Staw, 1995; Kacmar, 2009; Bedeian *et al.*, 2010; Tsui, 2013) observes, questionable research practices abound in Business and Management, too, a clear example being "data snooping or searching for asterisks (which) is the most damaging form of repeated testing, since the aim is to reject the null hypotheses while consciously ignoring the many models and tests that have been conducted and, thus, reporting greatly exaggerated levels of significance"[41]. Of course, the International Business, and Business and Management communities appear to recognize the issues at stake here. For instance, in an extended editorial in the *Academy of Management Journal* on this very issue, Dov Eden (2002, p. 841) noted that "a large number of high-quality replication studies" are needed to keep the discipline healthy. However, in Business and Management, too, really taking this badly needed medicine is easier said than done.

10. How to escape from this deadlock?

Looking at the editorial policies of the field's leading journals immediately gives the uneasy feeling that International Business is not that different. Take the following opening quote from the editorial policy of the field's top journal: "The *Journal of International Business Studies* (*JIBS*) is the top-ranked journal in the field of international business. The goal of *JIBS* is to publish insightful, innovative and impactful research on international business […] *JIBS* seeks to publish manuscripts with cutting-edge research that breaks new ground, rather than merely making an incremental contribution to international business studies" (webpage of the *Journal of International Business Studies*). In order to execute a coarse-grained check as to whether or not the International Business sub-discipline might need to reconsider current publication practices, given this initial observations, I downloaded all 2015 *Journal of International Business Studies* papers published to date (i.e. the January until August issues), and counted the number of reported negative, null, and positive results.

The findings are revealing[42]. Of the 32 published studies (excluding essay-type of articles, such as editorials and perspectives), 28 are of the hypotheses-testing kind, and not a single one is a replication study. In these 28 hypotheses-testing studies, a total of 142 hypotheses are tested. Of these 142 hypotheses, 124 are supported, which is 87 percent of the total. The number of nulls is 13 (9 percent), and the number of negatives is 5 (4 percent). Rather strikingly, in 20 of these 28 hypotheses-testing studies (71 percent), all hypotheses are supported. One outlier is an exception to the rule: in the paper of Levy *et al.* (2015), only 4 out of 12 hypotheses are supported – unnecessary to say that the above statistics will deteriorate substantially would this outlier be removed. These six authors, their study's reviewers, and the accepting Editor (David Thomas) are to be applauded. But one swallow does not make a summer.

So, what can we, as a collective of (International) Business and Management scholarship, do to escape from this dismal state of affairs?[43]. By and large, similar measures are suggested across a wide range of disciplines. In this essay, I would like to briefly discuss seven key measures (see e.g. Evanschitzky *et al.*, 2007; Mezias and Regnier, 2007 (in *Strategic Organization – déjà vu* indeed); Renkewitz *et al.*, 2011; Schooler, 2011):

(1) First, editorial policies might dispose of their current overly dominant pro-novelty and pro-positives biases, and explicitly encourage the publication of replication studies, including failed and unsuccessful ones that report null and negative findings. Apart from making this explicit in editorial policy statements, reviewers should be instructed along these lines, with appropriate reviewing forms and guidelines.

(2) Second, an option is to stimulate pre-reviewing/pre-publishing of a study's theory and design, as does *Comprehensive Results in Social Psychology*. Journals may either adopt a two-stage reviewing and publication process (theory and design first; evidence and discussion next), or open access registries of planned studies could be developed[44]. Essential is that, in so doing, scholars make their theory and predictions public before engaging in the empirical work. The empirical findings are guaranteed to be published, whether reflecting positives, nulls or negatives, provided the work is executed in a state-of-the-art manner.

(3) Third, open access publication by funding agencies and research institutes of all work produced prior to journal submission could provide access to studies not published in journals. This is essential for carrying out effective meta-analyses, avoiding the otherwise very difficult-to-correct publication bias. To the extent that this frustrates journals' double-blind reviewing processes, this is a price worth paying[45].

(4) Fourth, all raw data, protocols and data analysis codes[46] of accepted journal articles should be made available to the journal (which may collaborate with an established archive consortium) in order to make the execution of independent replication studies a way easier endeavor[47]. If there are good reasons for this, a (not too long) lag may be introduced before this material is made publicly available. Luckily, the Cloud's archival capacity is close to infinite.

(5) Fifth, a tradition of meta-analyses that correct for publication bias has to be established, similar to that in Medicine. To be able to run proper meta-analyses, the third and fourth measures are essential. After all, running meta-analyses with biased published results only is rather pointless, notwithstanding the option to apply modern correction techniques. Currently, finding non-published studies reporting nulls and negatives can be very challenging[48].

(6) Sixth, reporting significance only is inadequate, as the p-statistic is anything but uncontroversial (Nuzzo, 2014). Additionally, therefore, I would support Hubbard and Armstrong's (1997, p. 337) earlier plea for "reporting effect sizes and confidence intervals […] If statistical tests are used, power tests should accompany them." In this way, the overemphasis of significance is, at least partially, countered[49].

(7) Seventh, journals may appoint a replication section editor, as is done by, for example, the *Journal of Applied Econometrics*. This is a clear signal that the journal highly values this type of work, and indicates that space is available to publish well-executed and insightful replication studies. The latter can be achieved by, e.g., having a separate replications studies section, or by launching annual replication studies special issues.

Of course, by the end of the day, the above recommendations can only happen if there is a radical cultural change in the scientific community, including (International) Business and Management[50]. It should be in the community's DNA to engage in the quest for the "truth" – nothing more, nothing less. Such a change must involve all stakeholders: scholars, editors, reviewers, and students, but also funding agencies, research institutes, university presidents, faculty deans, department chairs, journalists, policymakers, and publishers. In the words of Ioannidis (2012, p. 647): "Safeguarding scientific principles is not something to be done once and for all. It is a challenge that needs to be met successfully on a daily basis both by single scientists and the whole scientific establishment."

Will all this happen automatically? Probably not, given vested interests and the notoriously difficult task of changing community cultures, if collective action is not effectively organized. In an attempt to organize such collective action, I opened a Pro-Falsification Petition Webpage: www.change.org/p/the-scientific-community-change-the-way-we-conduct-report-and-publish-our-research. This petition does not speak to the Business and Management community alone – quite the contrary. After all, the issues discussed in this essay are anything but new in many scientific (sub-)disciplines. And notwithstanding continued warnings by excellent scholars in all these (sub-)disciplines for many decades by now, the state of affairs appears to be changing from bad to worse (cf. Hubbard and Vetter, 1996; Evanschitzky *et al.*, 2007). Hence, all those who would support such a set of initiatives, inside and outside the (International) Business and Management community, are invited to post their thoughts and opinions on these important issues. And please pass on this link to as many colleagues as you can. In the meantime, we are discussing within the editorial team of *Cross-Cultural and Strategic Management* to see what we can do by leading by example. We will keep you posted.

Acknowledgment
Note that this essay only relates to issues intrinsic to the scientific community, which Tsui (2016) refers to as epistemic values. The author's thought-provoking essay provides a critical analysis of non-epistemic issues in the context of business schools, arguing that business school research is disconnected from practice and featuring a large pro-management bias. Basically, a growing unease can be observed between the practice-oriented focus of business school teaching *vis-à-vis* the science-obsessed orientation of Business and Management research (cf. Shapiro *et al.*, 2007; Corley and Gioia, 2011; Ghoshal, 2005; Sarasvathy, 2003; Starbuck, 2004; Tsui, 2013, 2015; Walsh *et al.*, 2003). In an attempt to counter the practice-science disconnect, *Public Administration Review* promotes the publication of reviews of scholarly articles by practitioners.This essay is based on a manifesto the author wrote in the Summer of 2015 (www.change.org/p/the-scientific-community-change-the-way-we-conduct-report-and-publish-our-research). After circulating an earlier draft of this manifesto among colleagues, the author received many responses. The author gratefully acknowledges the very constructive and supportive comments provided by Jean-Luc Arregle, Reinhard Bachmann, Julian Birkinshaw, Wouter van Bockhaven, Arjen Boin, Arjan van den Born, Bas Bosma, Steven Brakman, Boukje Cnossen, Per Davidsson, Gerald Davis, Marcus Dejardin, Desilava Dikova, Lorraine Eden, Karen Elliott, Marc Esteve, Dries Faems, Peer Fiss, Koen Frenken, César Garcia-Diaz, Vasiliki Gargalianou, Harry Garretsen, Richard Haans, Michael Hannan, Anne-Wil Harzing, Benson Honig,

Ann Jorissen, Wesley Kaufmann, Pim van Klink, Ruud Koning, Gerwin van der Laan, Joseph Lampel, Peter Leefland, Tim de Leeuw, Arie Lewin, Min Liu, Ellen Loots, Huigh van der Mandele, Katrin Mühlfeld, Giacomo Negro, Woody van Olffen, Kim Van Overvelt, Simon Parker, Gábor Péli, Jolien Philipsen, Padma Rao Sahib, Stephanie Rosenkranz, Saraï Sapulete, Joke Schrauwen, Jesse Segers, Tal Simons, Arndt Sorge, Linda Steg, David Storey, Doan Tru Trang, Anne Tsui, Rosalie Tung, Diemo Urbig, Johanna Vanderstraeten, Andy Van de Ven, Utz Weitzel, and Vicky Van Woensel. Of course, the usual disclaimer applies.

Notes

1. And a social community with the usual dose of politicking and power games: "the constitution of journals as 'top journals' is clearly an accomplishment of power. There is a circularity, in which to publish in the 'best' journals, one must produce the 'right kind' of work" (Grey, 2010, p. 683); and "It is argued that an effect of the 'one size fits all' logic of journal lists is to endorse and cultivate a research monoculture in which particular criteria, favored by a given list, assume the status of a universal benchmark of performance ('research quality')" (Mingers and Willmott, 2013, p. 1051). This practice discourages the pluralistic use of methods and perspectives that would provide the diversity essential for an effective and productive scientific "eco-system" (cf. Delbridge and Fiss, 2013).
2. Even without such biases, the literature will be full of false positives (and negatives) due to the nature of the significance rule of thumb that many of us apply routinely (cf. Mezias and Regnier, 2007). After all, would we all use a p-value threshold of 0.05, then 5 percent of the reported findings imply a Type I error. Hence, even if the scientific community would be without any human bias, replication is needed to filter out these stochastically generated Type I errors.
3. I try to do this, but many early-career scholars find that hard to accept, given the pressure they feel from the wider academic community.
4. Of course, Popper's falsifiability principle relates to theories, and not to each and every individual piece of scientific output. Actually, non-falsifiable research may well be valuable, too. This is true, for example, for initial conceptual theory-developing work, as well as for context- and history-specific case studies. Similarly, by way of mirror image, not all replication studies are valuable. Nulls and negatives may well be uninteresting (e.g. rejecting the hypothesis that organizations have green ears), or based on methodological quicksand (e.g. suffering from fatal endogeneity). Like any research, replication studies have to pass the hurdle of scientific scrutiny.
5. The Behavioral Economics Replication Project is a similar initiative launched in the Economics discipline, currently conducting a meta-experiment targeting 18 lab experiments published in 2011-2014 in the *American Economic Review* and *Quarterly Journal of Economics* (http://sciencepredictionmarkets.com/repoverview.html). Moreover, a few internet blogs and websites are actively reporting about scientific misconduct, emphasizing the need for replication studies – see, e.g., Retraction Watch (http://retractionwatch.com) and PubPeer (wwwpubpeer.com). Clearly, the issues discussed in this essay do not only affect the social sciences, but sciences at large.
6. This essay is directing the positivist scientific community engaged in quantitative research. Much of what is said below may well apply to non-positivist and non-quantitative research traditions too (see e.g. Golden, 1995), however, but that requires an essay of its own (see e.g. Sorge and Rothe, 2011). More generally, alternative epistemologies to Popper argue that either falsification should not be the main principle or that falsification only applies to a specific type of studies, and not to others – an argument that extends to the applicability

and value of replication studies (cf. Stengers, 2000). For instance, not all theory has to be falsifiable, and history-specific or context-bound case studies may not be replicable by their very idiosyncratic nature (see endnote 4). An example of a journal that explicitly takes an anti-positivist stance is *Organization* (see Parker and Thomas, 2011). A deep and engaging positivism – non-positivism dialogue (which currently tends to be, rather, a one-way monologue) might well be highly needed (Isaeva *et al.*, 2015), but integrating contrasting epistemologies is very difficult, and even impossible according to many (but see Schultz and Hatch, 1996).

7. The expansive definition of this label includes all Business and Management-related domains such as Accounting, Finance, Marketing and Operations. The more limited definition relates to the fields covered by the Academy of Management. In this essay, the main focus is on the latter, with quite a few trips to the former. Clearly, the arguments put forward here apply to International Business and Management as well. In Section 10, I briefly discuss the state of affairs in International Business and Management in particular. Much of what is argued applies to Business and Management broadly defined, and beyond – see for a recent example in accounting the Institute of Chartered Accountants of England and Wales (2015, pp. 17-18).
8. Of course, it is much simpler to compare means between two groups that are faced with different exogenous (experimental) treatments than to test for an effect in a model that is susceptible to model misspecification in the context of (non-experimental) noisy field data. Because effects are usually in terms of changes in the dependent variable due to a small change in the independent variable(s) of interest, and the dependent variable varies between studies (even though they tend to me more or less similar), it is harder to compare field studies and the effect sizes found. However, many meta-analytic studies using outcomes from field work have proven that this can be done.
9. The Big Data revolution may well push scientific practice from bad to worse. With thousands or millions of data points, not finding positives is very unlikely. Of course, in the data-mining literature, this danger is widely recognized, suggesting appropriate countermeasures (see e.g. Hand *et al.*, 2001).
10. Often, I receive(d) fatal comments by editors and reviewers of the following kind: “It is disappointing that you did not find any empirical support for your hypotheses,” and “ None of the hypotheses were supported. This is unfortunate, because the results are simply not interesting.” I am sure all of us do, provided that we dare to submit papers dominated by null and/or negative results.
11. Nothing new under the sun here. For instance, Norbert Wiener already complained about the inflation of published papers in the 1940s and 1950s, arguing that not creation (publication) is the problem, but selection – separating, in a huge pile, wheat from chaff.
12. Language is important, providing frames that can be productive or counterproductive. To really signal that such “incremental” work is needed, and that “groundbreaking” research feeds upon *ex ante* and benefits from *ex post* incremental studies, we may wish to replace the adjective “incremental” with “groundlaying”, to clearly reflect the synergetic relationship between the two. After all, building a wall requires many bricks.
13. Occasionally, we can. Gödel’s incompleteness theorems immediately shook the very basics of Mathematics in 1931, as did Einstein’s relativity theory with a few of Physics’ core assumptions in 1905 (a publication with zero references). In the social sciences, Coase’s (1937) transaction cost reasoning may be an example in Economics, as is Hannan and Freeman’s (1977) organizational ecology in Sociology.
14. *Strategic Organization* is yet another example of a journal struggling with this difficulty, apparently, of putting into practice what you preach. Notwithstanding two forceful pleas for

replication studies in this journal in 2007 (Helfat, 2007; Mezias and Regnier, 2007), an electronic search in their e-archive (2003-now) using "replication" as the search key in the title, abstract, and keywords generated zero hits beyond both pleas.

15. Strikingly Hubbard *et al.* (1998) find, for nine Business and Management journals in the 1976-1995 period, that replication studies are very scarce in both higher and lower-ranked outlets. So, the often-voiced claim that only top journals, with their heavy novelty bias, fail to publish replication studies cannot be supported. Moreover, the correlation between high impact and high quality is far from perfect.

16. These statistics are likely to improve substantially soon, as the *Strategic Management Journal* launched a special issue call titled "Replication in Strategic Management" (edited by Bettis, Helfat, and Shaver) in 2014. Recently, this was followed by a special issue call for meta-analysis proposals by the *Journal of Management Studies* (edited by Combs, Crook, and Rauch).

17. As Singh *et al.* (2003) point out, replication studies that contradict original work are even scarcer. Again, the use of specific language is significant, as such replication studies are frequently referred to as "failed" or "unsuccessful." Note that to the extent that (top) journals are willing to publish replication studies, they reveal a bias toward "successful" ones. To counter this in biomedicine, the *Journal of Negative Results in Biomedicine* only publishes "failed" replication studies (for other examples, see the *Journal of Errology* and PsychFileDrawer (http://psychfiledrawer.org/)).

18. An interesting recent launch is that of the *Journal of Business Venturing Insights*. Apart from offering a quick turnaround and space for heterodox contributions, this new outlet seeks to publish "non-findings or replication of established relationships" (see their website). A recent example of what this may produce is the replication and extension by Honig and Samuelsson (2014) of earlier work published by Delmar and Shane, and Delmar's (2015) response. Note that a similar plea for a *Journal of Replication Studies* was recently voiced in Economics (Zimmerman, 2015), while the *Political Science Replication Initiative* was launched in Political Science (http://projects.iq.harvard.edu/psreplication/data).

19. This is critical: only publishing replication studies with positives would simply aggravate the publication bias (Nuijten *et al.*, 2015).

20. Tsang and Kwan (1999) distinguish nine types, and even more can be thought of. For instance, conceptual replication implies that an original study is replicated with slightly different operationalizations of one or more of the key constructs. Yet another example is replication in the context of scale development. In the context of this essay, distinguishing three main types is sufficient.

21. Of course, the idea is not to publish such replications just for the sake of publishing replications. Such "literal" publication studies are probably only worth publishing in full if they contribute something to the original work – e.g., a methodological improvement (otherwise, a brief note will do, signaling a successful replication). To stimulate dialogue, the original authors may then be asked to post a reply.

22. Strictly speaking, this first type is closer to verification by re-doing the original analyses with the original data. In an ideal world, such control activities should not be necessary. However, given that the academic community is full of *homo sapiens* species featuring all common human biases, such an ideal world is utopian, implying that such control exercises are needed, both as an *ex post* check and an *ex ante* threat (see endnote 50).

23. Obviously, the study then has to be explicit about this, and carefully discuss the replication effort as a contribution in its own right. Often, baseline models (or "controls-only models") are skipped over without much reflection. Again, editors and reviewers stimulate this practice, by asking to focus on discussing the results related to the hypotheses.

24. Of course, this may reflect the practices of the day and/or the much smaller stock of prior work at the time, indicating that many taken-for-granted codes of conduct are period specific. Note, moreover, that Coase's (1937) essay was not recognized as path-breaking for decades. But, one could argue in a case like this, better late than never.
25. The cry for space for different types of work is anything but new (see e.g. Sutton and Staw, 1995). Some journals, including the *Journal of International Business Studies*, run a special section for freestyle-like essays. However, in practice, these essays tend to conform nicely to normal science practices. Moreover, a long list of journals has been established to promote "critical" perspectives (e.g. *Organization*, and *Critical Perspectives on International Business*). But, for one reason or the other, these alternative outlets are associated with non-positivist methodological editorial policies and preferences. A recent initiative is the *Academy of Management Discoveries*, "being open to replication studies. Replication studies provide a formalized way to detect anomalies by systematically investigating data in similar contexts and circumstances conform to a prior study" (Van de Ven *et al.*, 2015, p. 4). However, this new outlet focuses on empirical work only. Moreover, with the reference to the need for all contributions to be "rigorous", the door is open to launch all the usual urban myths and counterproductive practices. Notwithstanding these disclaimers, the number of new outlets trying to change current normal science practices is very encouraging indeed.
26. A broader issue relates to the counterproductive and low-quality nature of many of the current reviewing practices, implying too many Type I and Type II (non-)publication errors (i.e. publication of methodologically flawed studies and rejection of flawless papers). Quite a few colleagues who responded to my earlier request for comments on a first draft of the manifesto provided spine-chilling examples of their experiences with dysfunctional reviewing processes. Examples of such practices involve editors ignoring positive reviews, low-quality reviews of (probably) novice PhD students, pressure to cite the target journal and/or publications of the editor/reviewer, and rejection of manuscripts on the grounds of arguments that should have been raised at the desk reject stage. Three options to avoid some of the problems are to discard the double-blind reviewing practice, to thoroughly train reviewers (but see endnote 45), and/or to list the names of the reviewers after publication.
27. This is echoed in the recurring debates as to what good theory is, and whether Business and Management's theory fragmentation is good or bad (see e.g. Sutton and Staw in, 1995 *vis-à-vis* Corley and Gioia in 2011). Yet another alternative is to argue that Business and Management is a collection of normal science and non-normal science practices and theories: say, transaction cost economics and organizational ecology may be normal science paradigms (or sub-paradigms), but critical management studies and entrepreneurship might not. Mixing practices (and urban myths) from different root disciplines, such as Economics and Psychology, is anything but easy, if not impossible (cf. van Witteloostuijn, 2015). This essay is not the place to engage with this debate.
28. Often, these academic articles are a very difficult read, full of opaque jargon and complex argumentation. Manuscripts that are an easy read, written in common sense English, tend to be negatively evaluated by editors and reviewers, apparently assuming that such an easy read cannot be anything good.
29. This habit generates collateral damage by giving the impression to novices that we, as scholars, are able to immediately and without much effort produce such nicely polished end products. However, research is hard work, with all kinds of half-baked interim products along the way. Another example of collateral damage is bandwagon behavior, many starting to work on a "sexy" topic without sufficient attention for good theory, managerial relevance, and empirical rigor (see e.g. Lindebaum, 2015, on the neuroscience hype).
30. Interestingly, in disciplines such as Accounting, Finance, and Strategy, many teams chase for significant results in the very same secondary databases (often constructed and

provided by commercial enterprises, including WoS's Thomson Reuters). In replication repositories, these teams should specify their downloading procedures if the original data were under license so that other teams with licenses can download the exact same dataset.

31. Moreover, regrettably, by far the majority of replication studies in Business and Management, broadly defined, fail to correct for publication bias (by e.g. applying Duval and Tweedie's, 2000, trim-and-fill method), making their contribution to solving the issue very limited. Geyskens *et al.* (2009) evaluate 69 meta-analyses published in 14 Management top journals in 1980-2007 (of which only three in the *Journal of International Business Studies*), of which 57 (87 percent) do not correct for publication bias at all. The one meta-analysis I have been involved in myself so far (Bogaert *et al.*, 2016) does not apply a statistical publication bias-correcting technique either. However, this meta-analysis hints at another possible way out. Many of the ecological studies included in this meta-analysis add density and density squared – central variables in the meta-analysis – as control variables only, which is standard in organizational ecology's vital rates studies (i.e. regarding founding and mortality). My hunch is that reporting nulls and negatives for control variables is far more common than for their independent counterparts.
32. From a theoretical angle, Pfeffer and Fong (2005) argue in favor of developing organization theory from "first principles."
33. Much Business and Management research is conducted by scholars employed by commercial business schools, infusing Business and Management research with yet another channel of commercial influence (Tsui, 2016). Strikingly, many argue that this generates an overemphasis of the internal dynamic within the scientific community at the expense of the interests of external stakeholders, due to the dominance of all kinds of rankings (Adler and Harzing, 2009; Baum, 2011; Aguinis *et al.*, 2012), partly constructed on the basis of Thomson Reuters' statistics – a peculiar feedback loop indeed. A few initiatives have been launched in an attempt to deconstruct this damaging feedback loop inside and outside Business and Management (see e.g. the "Socially Responsible Scholarship" or "Responsible Science" grassroots movement in the Academy of Management, and the San Francisco Declaration of Research Assessment 2012: cf. Tsui, 2013, 2015, 2016; https://en.wikipedia.org/wiki/San_Francisco_Declaration_on_Research_Assessment).
34. That is, officially, Web of Science® Journal Citation Reports® 2014 Edition.
35. But the PLoS initiative does offer a solution for a series of other issues, such as the nature of the reviewing process and open access availability. However, not all open access initiatives are equally open, as quite a few come with high publication or submission fees, including *PLoS ONE* (for another example, see e.g. *Sociological Science*).
36. This implies the need to seriously re-think current practices. For instance, the criteria for tenure and promotion decisions are too often overly dominated by quantitative publication criteria, emphasizing publication in a few journals with a high Impact Factor (®!). Another option is that schools may establish "Replication Chairs."
37. This offers a brilliant opportunity to engage in the usual obligatory referencing to classic studies: DiMaggio and Powell (1983) institutional theory of population-level isomorphism, and Hannan and Freeman's (1984) organizational ecology of organization-level inertia. Both perspectives provide ample arguments as to why changing the world-wide academic community or any individual academic institution is close to a mission impossible (cf. endnote 50).
38. Not only replication work suffers from low status. Another example is descriptive and data-driven research. Helfat (2007) argues that Business and Management should encourage studies that report "stylized facts", which is seen as important groundlaying work in many other disciplines (take the case of Biology, in which Darwin could not have done his groundbreaking work without knowledge of many "stylized facts").

39. Doing only this would be problematic, reinforcing the low-status reputation of replication work.

40. Simmons *et al.* (2011) refer to researcher degrees of freedom as the root of many of these issues. That is, behind closed doors, researchers make a whole series of decisions that affect the reported results, from sample size and outlier treatment to selection of control variables and measurement transformation: "it is common (and accepted practice) for researchers to explore various analytic alternatives, to search for a combination that yields 'statistical significance' " (Simmons *et al.*, 2011, p. 1359). These practices, of course, further boost the (false) positives bias. They suggest to introduce six reporting requirements, which have to be carefully checked by reviewers, to bring these hidden decisions into the open. Note that Head *et al.* (2015) argue that the impact of such so-called *p*-hacking on reported effect sizes is rather weak (but see Brodeur *et al.*, 2012).

41. This squarely runs against the AOM Code of Ethics (http://aom.org/About-AOM/Code-of-Ethics.aspx), which indicates that such codes, however well meant, tend to be rather ineffective.

42. This counting exercise was not always easy, as the transparency of reporting is rather uneven. Moreover, for the sake of parsimony, I here ignore the levels of significance and the effect sizes. For the purpose of this essay, the numbers reported in the main text suffice. And of course, a rigorous counting exercise requires input from multiple coders.

43. This offers yet another opportunity to add yet another reference to a social sciences classic: Olson's (1965) *The Logic of Collective Action*. Organizing collective action is notoriously difficult in a heterogeneous community full of people and institutions with conflicting interests, and motivated by perverse incentives.

44. Strictly speaking, *Comprehensive Results in Social Psychology* expects that the first step of the reviewing process is entered before the empirical research is conducted (quite a gaming-sensitive expectation). By way of experiment, the *Journal of Business Psychology* has recently introduced an alternative two-step reviewing process for empirical work that has already been carried out, implying that the introduction-theory-design part of the paper can be submitted first after the empirical work has been performed. For a similar initiative in Economics, see the AEA RCT Registry (www.socialscienceregistry.org/), which is mainly used for field experiments (and not for non-experimental field work). Early January 2016, ten journals in the HR and organizational behavior domain announced in concert that they would experiment with a hybrid registered reports submission path, including *Leadership Quarterly* and *Organizational Research Methods*.

45. Already now, in the wonderful world of the World-Wide Web, reviewers can, if they want to for whatever peculiar reason, by-pass the double-blind requirement. Regrettably, this may imply that established authors will have an easier time while early-career scholars with no publication record will have a tougher time to break into publishing in good journals. A remedy may be to seriously train reviewers (but this is anything but easy: see Callaham and Tercier, 2007; Schroter *et al.*, 2008). Another complication is that many journals explicitly refuse to review papers that circulate in other forms – a policy that is routinely circumvented by authors anyway (e.g. by changing the title of earlier working papers before submission).

46. Or any other relevant material. For instance, for computer simulation studies, this should involve software programming codes (cf. Delbridge and Fiss, 2013).

47. Examples of journals requiring this are the *American Economic Review*, *Journal of Conflict Resolution*, *Journal of Money*, *Credit and Banking*, and *American Journal of Political Science*. For instance, the *American Economic Review* has a one-pager on its data availability policy, starting with the strict statement that "It is the policy of the *American Economic Review* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication" (from their webpage; emphasis added). Regrettably, this does not imply that the material posted is guaranteed to be sufficient for replication studies – quite the contrary (cf. Evanschitzky *et al.*, 2007).

48. As said, another issue is that, in all likelihood, non-published studies suffer from the positives bias, too. This implies that the suggested measures must be taken in concert.
49. Note that, of course, *post hoc* power tests regarding null findings would be counterproductive (Hoenig and Heisey, 2001). More broadly, many reporting urban myths have to be tackled. Another one, for instance, relates to using p-values only to test null hypotheses, rather than β coefficients or non-null hypotheses. More generally, not all studies are fully up-to-date regarding statistical methods. In the statistics literature, much research has been done into estimation and testing of causal effects (see e.g. the recent book by Imbens and Rubin, 2015). For instance, given the often small sample sizes in (International) Business and Management, one would expect tests to be validated using bootstrap methods, so one does not rely on asymptotic results.
50. Much of the above is simply blocked from within the scientific community. For instance, Wicherts *et al.* (2011) conclude that psychologists are often unwilling or unable to share their data for re-analysis, which is unlikely to be very different in (International) Business and Management (cf., in Economics, Dewald *et al.*, 1986; McCullough *et al.*, 2006, 2008). We as (International) Business and Management scholars are experts in the study of organizational change; hence, this unwillingness to engage in radical change does not come as a surprise (Hannan and Freeman, 1984). And changing a community of organizations is even harder, as we know from institutional theory (DiMaggio and Powell, 1983), especially when the individual-level incentives are anti-change and collective action is needed (Olson, 1965).

References

Adler, N.J. and Harzing, A.-W. (2009), "When knowledge wins: transcending the sense and nonsense of academic rankings", *Academy of Management Learning & Education*, Vol. 8, pp. 72-95.

Aguinis, H.A., Suarez-Gonzalez, I., Lannelongue, G. and Joo, H. (2012), "Scholarly impact revisited", *Academy of Management Perspectives*, Vol. 26, pp. 105-132.

Barker, V.L. and Mone, M.A. (1994), "Retrenchment: cause of turnaround or consequence of decline?", *Strategic Management Journal*, Vol. 15, pp. 395-405.

Baum, J.A. (2011), "Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies", *Organization*, Vol. 18, pp. 449-466.

Bedeian, A.G. (2003), "The manuscript review process: the proper roles of authors, referees, and editors", *Journal of Management Inquiry*, Vol. 12, pp. 331-338.

Bedeian, A.G., Taylor, S.G. and Miller, A.N. (2010), "Management science on the credibility bubble: cardinal sins and various misdemeanors", *Academy of Management Learning & Education*, Vol. 9, pp. 715-725.

Begg, C.B. and Berlin, J.A. (1988), "Publication bias: a problem in interpreting medical data", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 151, pp. 419-463.

Bettis, R.A. (2012), "The search for asterisks: comprised statistical tests and flawed theories", *Strategic Management Journal*, Vol. 33, pp. 108-113.

Birkinshaw, J., Healey, M.P., Suddaby, R. and Weber, K. (2014), "Debating the future of management research", *Journal of Management Studies*, Vol. 51, pp. 38-55.

Blaug, M. (1992), *The Methodology of Economics: Or How Economists Explain*, Cambridge University Press, New York, NY.

Bogaert, S., Boone, C., Negro, G. and van Witteloostuijn, A. (2016), "Organizational form emergence: a meta-analysis of the ecological theory of legitimation", *Journal of Management*, Vol. 42, pp. 1344-1373.

Bradford De Long, J. and Lang, K. (1992), "Are all economic hypotheses false?", *Journal of Political Economy*, Vol. 100, pp. 1257-1272.

Brodeur, A., Lé, M., Sangnier, M. and Zylberberg, Y. (2012), "Star wars: the empirics strike back", Working Paper No. 2012-29, Paris School of Economics, Paris.

Callaham, M. and McCulloch, C. (2011), "Longitudinal trends in the performance of scientific peer reviewers", *Annals of Emergency Medicine*, Vol. 57, pp. 141-148.

Callaham, M.L. and Tercier, J. (2007), "The relationship of previous training and experience of journal peer reviewers to subsequent review quality", *PLoS Medicine*, Vol. 4, pp. 32-40.

Card, D. and Krueger, A.B. (1995), "Time-series minimum wage studies: a meta-analysis", *American Economic Review Papers and Proceedings*, Vol. 85, pp. 238-243.

Coase, R.H. (1937), "The nature of the firm", *Economica*, Vol. 16, pp. 386-405.

Colquitt, J.A. and Zapata-Phelan, C.P. (2007), "Trends in theory building and theory testing: a five-decade study of the *Academy of Management Journal", Academy of Management Journal*, Vol. 50, pp. 1281-1303.

Corley, K.G. and Gioia, D.A. (2011), "Building theory about theory building: what constitutes a theoretical contribution?", *Academy of Management Review*, Vol. 36, pp. 12-32.

Davis, G.F. (2015), "Editorial essay: what is organizational research for?", *Administrative Science Quarterly*, Vol. 60, pp. 179-188.

Delmar, F. (2015), "A response to Honig and Samuelson (2014)", *Journal of Business Venturing Insights*, Vol. 3, pp. 1-4.

Delbridge, R. and Fiss, P.C. (2013), "Editors' comments: styles of theorizing and the social organization of knowledge", *Academy of Management Review*, Vol. 38, pp. 325-331.

Dewald, W.G., Thursby, J.G. and Anderson, R.G. (1986), "Replication in empirical economics: the journal of money, credit and banking project", *American Economic Review*, Vol. 76, pp. 587-603.

Dikova, D. and van Witteloostuijn, A. (2007), "Acquisition versus greenfield foreign entry: diversification mode choice in Central and Eastern Europe", *Journal of International Business Studies*, Vol. 38, pp. 1013-1033.

DiMaggio, P. and Powell, W.W. (1983), "The iron cage revisited: collective rationality and institutional isomorphism in organizational fields", *American Sociological Review*, Vol. 48, pp. 147-160.

Doucouliagos, H. and Stanley, T.D. (2009), "Publication bias in minimum-wage research? A meta-regression analysis", *British Journal of Industrial Relations*, Vol. 47, pp. 406-428.

Duval, S. and Tweedie, R. (2000), "A nonparametric 'Trim and Fill' method of accounting for publication bias in meta-analysis", *Journal of American Statistical Association*, Vol. 95, pp. 89-98.

Eden, D. (2002), "Replication, meta-analysis, scientific progress, and *AMJ*'s publication policy", *Academy of Management Journal*, Vol. 45, pp. 841-846.

Evanschitzky, H., Baumgarth, C., Hubbard, R. and Armstrong, J.S. (2007), "Replication research's disturbing trend", *Journal of Business Research*, Vol. 60, pp. 411-415.

Ferguson, C.J. and Heene, M. (2012), "A vast graveyard of undead theories: publication bias and psychological science's aversion to the null", *Perspectives on Psychological Science*, Vol. 7, pp. 555-561.

Geyskens, I., Krishnan, R., Steenkamp, J.-B.E.M. and Cunha, P.V. (2009), "A review and evaluation of meta-analysis practices in management research", *Journal of Management*, Vol. 35, pp. 393-419.

Ghoshal, S. (2005), "Bad management theories are destroying good management practices", *Academy of Management Learning & Education*, Vol. 4, pp. 75-91.

Golden, M.A. (1995), "Replication and non-quantitative research", *PS: Political Science and Politics*, Vol. 28, pp. 481-483.

Grey, C. (2010), "Organizing studies: publications, politics and polemic", *Organization Studies*, Vol. 31, pp. 677-694.

Hambrick, D.C. (2007), "The field of management's devotion to theory: too much of a good thing?", *Academy of Management Journal*, Vol. 50, pp. 1346-1352.

Hand, D., Mannila, H. and Smyth, P. (2001), *Principles of Data Mining*, MIT Press, Cambridge, MA.

Hannan, M.T. and Freeman, J. (1977), "The population ecology of organizations", *American Journal of Sociology*, Vol. 82, pp. 929-964.

Hannan, M.T. and Freeman, J. (1984), "Structural inertia and organizational change", *American Sociological Review*, Vol. 49, pp. 149-164.

Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. and Jennions, M.D. (2015), "The extent and consequences of P-hacking in science", *PLoS Biology*, Vol. 13, e10022106.

Helfat, C.E. (2007), "Stylized facts, empirical research and theory development in management", *Strategic Organization*, Vol. 5, pp. 185-192.

Hoenig, J.M. and Heisey, D.M. (2001), "The abuse of power: the pervasive fallacy of power calculations for data analysis", *The American Statistician*, Vol. 55, pp. 1-6.

Honig, B. and Samuelsson, M. (2014), "Data replication and extension: a study of business planning and venture-level performance", *Journal of Business Venturing Insights*, Vols 1-2, pp. 18-25.

Hubbard, R. and Armstrong, J.S. (1997), "Publication bias against null results", *Psychological Reports*, Vol. 80, pp. 337-338.

Hubbard, R. and Vetter, D.E. (1996), "An empirical comparison of published replication research in accounting, economics, finance, management, and marketing", *Journal of Business Research*, Vol. 35, pp. 153-164.

Hubbard, R., Vetter, D.E. and Little, E.L. (1998), "Replication in strategic management: scientific testing for validity, generalizability, and usefulness", *Strategic Management Journal*, Vol. 19, pp. 243-254.

Imbens, G.W. and Rubin, D.B.D.B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, Cambridge.

Institute of Chartered Accountants in England and Wales (2015), *The Effects of Mandatory IFRS Adoption in the EU: A Review of Empirical Research*, ICAEW, London.

Ioannidis, J.P.A. (2005), "Why most published research findings are false", *PLoS Medicine*, Vol. 2, e124.

Ioannidis, J.P.A. (2012), "Why science is not necessarily self-correcting", *Perspectives on Psychological Science*, Vol. 7, pp. 645-654.

Isaeva, N., Bachmann, R., Bristow, A. and Saunders, M.N.K. (2015), "Why the epistemologies of trust researchers matter", *Journal of Trust Research*, forthcoming.

Jefferson, T., Alderson, P., Wagner, E. and Davidoff, F. (2002), "Effects of editorial peer review: a systematic review", *Journal of the American Medical Association*, Vol. 287 No. 21, pp. 1-4.

Jonas, K.J. and Cesario, J. (2015), "How can preregistration contribute to research in our field?", *Comprehensive Results in Social Psychology*. doi: 10.1080/23743603.2015.1070611.

Kacmar, K.M. (2009), "From the editors: an ethical quiz", *Academy of Management Journal*, Vol. 52, pp. 432-434.

Kahneman, D. (2011), *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, NY.

Kane, E.J. (1984), "Why journal editors should encourage the replication of applied econometrics research", *Quarterly Journal of Business & Economics*, Vol. 23, pp. 3-8.

Kerr, N.L. (1998), "HARKing: hypothesizing after the results are known", *Personality and Social Psychology*, Vol. 2, pp. 196-217.

Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, IL.

Leamer, E.E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, New York, NY.

Levy, O., Taylor, S., Boyacigiller, N.A., Bodner, T.E., Peiperl, M.A. and Beechler, S. (2015), "Perceived senior leadership opportunities in MNCs: the effect of social hierarchy and capital", *Journal of International Business Studies*, Vol. 46, pp. 285-307.

Lexchin, J., Bero, L.A., Djulbegovic, B. and Clark, O. (2003), "Pharmaceutical industry sponsorship and research outcome and quality: systematic review", *British Medical Journal*, Vol. 326, pp. 1167-1170.

Lindebaum, D. (2015), "Critical essay: building new management theories on sound data? The case of neuroscience", *Human Relations*, forthcoming.

McCullough, B.D., McGeary, K.A. and Harrison, T.D. (2006), "Lessons from the *JMCB* archive", *Journal of Money, Credit and Banking*, Vol. 38, pp. 1093-1107.

McCullough, B.D., McGeary, K.A. and Harrison, T.D. (2008), "Do economics journal archives promote replicable research?", *Canadian Journal of Economics*, Vol. 41, pp. 1406-1420.

Makel, M., Plucker, J. and Hegarty, B. (2012), "Replications in psychology research: how often do they really occur?", *Perspectives in Psychological Science*, Vol. 7, pp. 537-542.

Mayer, M. and Whittington, R. (2003), "Diversification in context: a cross-national and cross-temporal extension", *Strategic Management Journal*, Vol. 24, pp. 773-781.

Melander, H., Ahlqvist-Rastad, J., Meijer, G. and Beermann, B. (2003), "Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications", *British Medical Journal*, Vol. 326, pp. 1171-1173.

Mezias, S.J. and Regnier, M.O. (2007), "Walking the walk as well as talking the talk: replication and the normal science paradigm in strategic management research", *Strategic Organization*, Vol. 5, pp. 283-296.

Miller, D. (2007), "Paradigm prisons, or in praise of atheoretic research", *Strategic Organization*, Vol. 5, pp. 177-184.

Million, M. and Raoult, D. (2012), "Publication biases in probiotics", *European Journal of Epidemiology*, Vol. 27, pp. 885-886.

Mingers, J. and Willmott, H. (2013), "Taylorizing business school research: on the "one best way" performative effects of journal ranking lists", *Human Relations*, Vol. 66, pp. 1051-1073.

Mittelstaedt, R.A. and Zorn, T.S. (1984), "Economic replications: lessons from the experimental sciences", *Quarterly Journal of Business and Economics*, Vol. 23, pp. 9-15.

Moreno, S.G., Sutton, A.J., Ades, A.E., Cooper, N.J. and Abrams, K.R. (2011), "Adjusting for publication biases across similar interventions performed well when compared with gold standard data", *Journal of Clinical Epidemiology*, Vol. 64, pp. 1230-1241.

Moreno, S.G., Sutton, A.J., Turner, E.H., Abrams, K.R., Cooper, N.J., Palmer, T.M. and Ades, A.E. (2009), "Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications", *British Medical Journal*, Vol. 339, b2981.

Neuroskeptic (2012), "The nine circles of scientific hell", *Perspectives on Psychological Science*, Vol. 7, pp. 643-644.

Nuijten, M.B., van Assen, M.A.L.M., Veldkamp, C.L.S. and Wicherts, J.M. (2015), "The replication paradox: combining studies can decrease accuracy of effect size estimates", *Review of General Psychology*, forthcoming.

Nuzzo, R. (2014), "P Values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume", *Nature*, available at: www.nature.com/news

Olson, M. (1965), *The Logic of Collective Action*, Harvard University Press, Cambridge, MA.

Open Science Collaboration (2015), "Estimating the reproducibility of psychological science", *Science*, Vol. 349, aac4716-1.

Parker, M. and Thomas, R. (2011), "What is a critical journal?", *Organization*, Vol. 18, pp. 419-427.

Pashler, H. and Wagenmakers, E.-J. (2012), "Editors' introduction to the special section on replicability in psychological science: a crisis of confidence?", *Perspectives on Psychological Science*, Vol. 7, pp. 528-530.

Pfeffer, J. (1993), "Barriers to the advance of organizational science: paradigm development as a dependent variable", *Academy of Management Review*, Vol. 18, pp. 599-620.

Pfeffer, J. and Fong, C.T. (2005), "Building organization theory from first principles", *Organization Science*, Vol. 16, pp. 372-388.

Popper, K. (1959), *The Logic of Scientific Discovery*, Routledge, Oxford.

Renkewitz, F., Fuchs, H.M. and Fiedler, S. (2011), "Is there evidence of publication bias in *JDM* research?", *Judgement and Decision Making*, Vol. 6, pp. 870-881.

Rosenthal, R. (1979), "An introduction to the file drawer problem", *Psychological Bulletin*, Vol. 86, pp. 638-641.

Sarasvathy, S. (2003), "Entrepreneurship as a science of the artificial", *Journal of Economic Psychology*, Vol. 24, pp. 203-220.

Schooler, J. (2011), "Unpublished results hide the decline effect", *Nature*, Vol. 470 p. 437.

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L. and Smith, R. (2008), "What errors do peer reviewers detect, and does training improve their ability to detect them?", *Journal of the Royal Society of Medicine*, Vol. 101, pp. 507-514.

Schultz, M. and Hatch, M.J. (1996), "Living with multiple paradigms: the case of paradigm interplay in organizational culture studies", *Academy of Management Review*, Vol. 21, pp. 529-557.

Shapiro, D.L., Kirkman, R.L. and Courtney, H.G. (2007), "Perceived causes and solution of the translation problem in management research", *Academy of Management Journal*, Vol. 50, pp. 249-266.

Shleifer, A. (2012), "Psychologists at the gate: a review of Daniel Kahneman's thinking, fast and slow", *Journal of Economic Literature*, Vol. 50, pp. 1080-1091.

Simmons, J.P., Nelson, L.D. and Simonsohn, U. (2011), "False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant", *Psychological Science*, Vol. 22, pp. 1359-1366.

Singh, K., Ang, S.H. and Leong, S.M. (2003), "Increasing replication for knowledge accumulation in strategy research", *Journal of Management*, Vol. 29, pp. 533-549.

Sorge, A.M. and Rothe, K. (2011), "Resource dependence and construction, and macro and micro politics in transnational enterprises and alliances: the case of jet engine manufacturers in Germany", in Geppert, M. and Dörrenbächer, C. (Eds), *Politics and Power in the Multinational Corporation: The Role of Institutions, Interests and Identities*, Cambridge University Press, Cambridge, pp. 41-71.

Stanley, T.D. (2001), "Wheat from chaff: meta-analysis as quantitative literature review", *Journal of Economic Perspectives*, Vol. 15, pp. 131-150.

Stanley, T.D. (2008), "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection", *Oxford Bulletin of Economics and Statistics*, Vol. 70, pp. 103-127.

Starbuck, W.H. (2003), "Turning lemons into lemonade: where is the value in peer reviews?", *Journal of Management Inquiry*, Vol. 12, pp. 344-351.

Starbuck, W.H. (2004), "Why I stopped trying to understand the real world", *Organization Studies*, Vol. 25, pp. 1233-1254.

Stengers, I. (2000), *The Invention of Modern Science*, Vol. 19, University of Minnesota Press, Minneapolis, MI.

Sterling, T.D. (1959), "Publication decisions and their possible effects on inferences drawn from tests of significance", *Journal of the American Statistical Association*, Vol. 54, pp. 30-34.

Sterne, J.A.C., Egger, M. and Smith, G.D. (2001), "Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis", *British Medical Journal*, Vol. 323, pp. 101-105.

Sutton, R.I. and Staw, B.M. (1995), "What theory is not", *Administrative Science Quarterly*, Vol. 40, pp. 371-384.

The Economist (2013), "Trouble at the lab", *The Economist*, available at: www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble (accessed July 30, 2015).

Tsang, E.W.K. and Kwan, K.M. (1999), "Replication and theory development in organizational science: a critical realist perspective", *Academy of Management Review*, Vol. 24, pp. 759-780.

Tsui, A.S. (2013), "On compassion in scholarship: why should we care?", *Academy of Management Review*, Vol. 38, pp. 167-181.

Tsui, A.S. (2015), "Reconnecting with the business world: socially responsible scholarship", *EFMD Global Focus*, Vol. 9 No. 1, pp. 36-39.

Tsui, A.S. (2016), "Reflections on the so-called value-free ideal: values and science in the business schools", *Cross-Cultural and Strategic Management (* formerly known as *Cross Cultural Management)*, Vol. 23, pp. 4-28.

Tsui, A.S. and Hollenbeck, J.R. (2009), "Successful authors and effective reviewers balancing supply and demand in the organizational sciences", *Organizational Research Methods*, Vol. 12, pp. 259-275.

Van de Ven, A.H., Ang, S., Arino, A., Bamberger, P., LeBaron, C., Miller, C. and Milliken, F. (2015), "Welcome to the *Academy of Management Discoveries* (*AMD*)", *Academy of Management Discoveries*, Vol. 1, pp. 1-4.

Van Maanen, J. (1989), "Some notes on the importance of writing in organization studies", Harvard Business School Research Colloquium, Harvard Business School, Cambridge, MA, pp. 27-33.

van Witteloostuijn, A. (2015), "Toward experimental international business: unraveling fundamental causal linkages", *Cross-Cultural and Strategic Management (*formerly known as *Cross Cultural Management)*, Vol. 22, pp. 530-544.

Walsh, J.P., Weber, K. and Margolis, J.D. (2003), "Social issues and management: our lost cause found", *Journal of Management*, Vol. 29, pp. 859-881.

Wicherts, J.M., Bakker, M. and Molenaar, D. (2011), "Willingness to share research data is related to the strength of the evidence and the quality of the reporting of statistical results", *PLoS ONE*, Vol. 6, e26828.

Wigboldus, D.H.J. and Dotsch, R. (2015), “Encouraging playing with data and discourage questionable reporting practices”, *Psychometrika*, forthcoming.

Yong, E. (2013), “Psychologists strike a blow for reproducibility: thirty-six labs collaborate to check 13 earlier findings”, *Nature News*, available at: www.nature.com/news/

Zimmerman, C. (2015), “On the need for a replication journal”, Working Paper No. 2015-016A, Federal Reserve Bank of St. Louis, St. Louis, MO.

Corresponding author
Arjen van Witteloostuijn can be contacted at: A.vanWitteloostuijn@uvt.nl

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

This article has been cited by:

1. HarzingAnne-Wil Anne-Wil Harzing Anne-Wil Harzing (PhD, Bradford) is a Professor of International Management at Middlesex University, UK. Her research interests include international HRM, HQ-subsidiary relationships, the role of language in international business, and the quality and impact of academic research. She has published nearly 100 refereed journal articles and books/book chapters, and has been listed on Thomson Reuter's Essential Science Indicators top 1 percent most cited academics in Economics & Business worldwide since 2007. Since 1999 she maintains an extensive website (www.harzing.com) with resources for academic publishing, including the Journal Quality List and Publish or Perish, a software program that retrieves and analyzes academic citations. Business School, Middlesex University, London, UK . 2016. Why replication studies are essential: learning from failure and success. *Cross Cultural & Strategic Management* **23**:4, 563-568. [Abstract] [Full Text] [PDF]