What happened to Popperian Falsification? A Manifesto to Create a Healthier Business and Management Scholarship — Toward a Scientific Wikipedia

Arjen van Witteloostuijn [1] [2] [3]

September 2015

[1] Tilburg University, The Netherlands
[2] University of Antwerp and Antwerp Management School, Belgium
[3] Cardiff Business School, United Kingdom

Abstract. Current publication practices in the scholarly 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. Indeed, 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, Business and Management research is ending up in a deadlock. In this Manifesto, I extensively argue what I believe is wrong, why that is so, and what we might do about this.

Keywords. Falsification principle; verification bias; false positives; publication practices.

Acknowledgements

After circulating an earlier draft of this Manifesto among colleagues, I received many responses. I gratefully acknowledge 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, Gerald Davis, Marcus Dejardin, Dessi 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, Ann Jorissen, Wesley Kaufmann, Pim van Klink, Ruud Koning, Gerwin van der Laan, Peter Leeflang, Tim de Leeuw, 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. Bas Bosma's programming skills produced this very professional layout. The usual disclaimer applies.

A Manifesto

THIS IS AN ACADEMIC MANIFESTO,[4] implying a peculiar mixture of polemical, political and scholarly arguments.[5] This Manifesto is primarily, but not exclusively, directed at the worldwide Business and Management scholarly community, because I passionately believe that our discipline, or set of disciplines, must deal with a number of deeply-rooted problems, which can be summarized as the publication bias and replication defect crisis: we, as a collective, violate very basic scientific principles (a) by mainly publishing positive findings (i.e., those that are in support of our hypotheses) and (b) by rarely engaging in replication studies (being obsessed with preferrably "cutting-edge" and "groundbreaking" novelty). Clearly, Business and Management is not the only discipline in crisis — quite the contrary. But the least we can do is to try to clean up our own mess. Below, I rather extensively argue what I believe is wrong, why that is so, and what we might do about this. I do so not only in the main text, but also in a lengthy series of remarks in side notes in the margin.

Of course, I do not claim to be the beholder of all truth. For sure, this Manifesto is incomplete; for sure, this Manifesto hosts mistakes; and for sure, you will not agree with everything being said. However, if you share the worries, by and large, expressed in this Manifesto, I would highly appreciate if you could explicitly signal your support. For that purpose, I opened a Pro-Falsification Petition Webpage that can be signed, and which can be used to start exchanging ideas: `https://www.change.org`. To kick-start this dialogue, I provide a (very) brief and tentative suggestion regarding a new way of publishing, for now referred to as Scientific Wikipedia, in the Appendix. My hope is that by initiating this dialogue, a few of the measures suggested below will indeed be implemented, and others — perhaps far more effective ones — will be added in due course. The time is right to start organizing collective action to improve the state of our beautiful and wonderful Business and Management scientific community.

[4] This Manifesto grew out of a far more modest editorial I wrote for *Cross-Cultural and Strategic Management* (2016), which explains the special attention for International Business. Note that this Manifesto only relates to issues intrinsic to the scientific community, which Tsui (2016) refers to as epistemic values. Her thought-provoking essay provides a critical analysis of non-epistemic issues in the context of business schools, arguing that business school research is (a) disconnected from practice and (b) 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. Corley and Gioia, 2011; Ghoshal, 2005; Sarasvathy, 2003; Shapiro et al., 2007; 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.

[5] "A *manifesto* is a published verbal declaration of the intentions, motives, or views of the issuer, be it an individual, group, political party or government. A manifesto usually accepts a previously published opinion or public consensus and/or promotes a new idea with prescriptive notions for carrying out changes the author believes should be made. It often is political or artistic in nature, but may present an individual's life stance" (Wikipedia). Much of this definition applies to this Manifesto as well. However, being an academic, I could not resist to mix this Manifesto format with scholarly elements, such as many remarks in (margin) notes and 100-plus(!) references.

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.[6] This implies that science as a community is subject to all the same behavioral patterns that all human communities are, including a plethora of 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.[7] 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.[8] 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, but not only there, ample evidence abounds that current reviewing practices fail to provide the effective filtering mechanism they are claimed to provide (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 McCulloch, 2011; 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 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

[6] 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, 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, favoured by a given list, assume the status of a universal benchmark of performance ('research quality')" (Mingers and Willmott, 2013, 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" (Delbridge and Fiss, 2013).

[7] 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 .05, then 5 per cent 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.

[8] For a critical discussion regarding Business and Management, see, e.g., Bedeian (2003); Starbuck (2003); Tsui and Hollenbeck (2009).

teach the younger generation of researchers that instead of being overly discouraged, they should be happy if they can*not* confirm their hypotheses. This quest for falsification is critical because, in the words of Ioannidis (2012, 646), "Efficient and unbiased replication mechanisms are essential for maintaining high levels of scientific credibility."[9] 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.[10]

[9] 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.

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 (Lexchin et al., 2003; Melander et al., 2003): (false) positives are published, whilst nulls and negatives are not. The recent AllTrials initiative (`http://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.

[10] 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). So, like any research, replication studies have to pass the hurdle of scientific scrutiny.

In Psychology, a discipline closer to Business and Management,[11] 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* (`http://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, September 26 2012).

[11] 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 Manifesto, the main focus is on the latter, with quite a few trips to the former. However, 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 in England and Wales (2015, 17–18).

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 per cent of positives, only 36 per cent 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).[12] Third, in 2015, the tailor-made *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" (`http://www.tandfonline.com/loi/rrsp20#.VbtCRDcw-70`; see also the launch Editorial by Jonas and Cesario (2015)).

[12] 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 (`http://wwwpubpeer.com`). Clearly, the issues discussed in this Manifesto do not only affect the Social Sciences, but Sciences at large.

[13] 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).

The Verification Bias

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.[13] And related to this, the countermeasures involve the registration and/or replication of such lab and/or trial work. This Manifesto is primarily directing the positivist Business and Management scientific community engaged in quantitative research.[14] 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. 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

[14] Much of what is said below may well apply to non-positivist and non-quantitative research traditions too (Golden, 1995), however, but that requires a Manifesto 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 margin note 10).

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 be 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. Certainly, at its 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 at all. There appears to be a disconnect between what we practice from what we preach.[15] In reality, we are obsessively focused 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 whilst 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.[16] 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.[17]

Faced with a "publish or perish" culture, 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-

[15] 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.

[16] 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).

[17] 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.

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.

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 — how boring that is.

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[18] 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

[18] 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.

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[19] 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 ecosystems 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.

[19] 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 transaction cost reasoning (1937) may be an example in Economics, as is Hannan and Freeman's organizational ecology (1977) in Sociology.

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. 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, Hannan and 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, 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*.[20] 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).[21] 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).[22] 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 and 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."

In light of the above, the time is right for launching a *Journal of Business and Management Replication Studies*. Without replication studies, successful and failed ones, Business and Management will end up in a deadlock. This is critical: only publishing replication studies with positives would simply aggravate the publication bias (Nuijten et al., 2015). 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 Manifesto, distinguishing three main types is sufficient. I would suggest to promote at least these three types of replication studies (cf. Tsang and Kwan, 1999).

The first type is that other scholars replicate a published study's analyses with this study's data.[23] 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.[24] 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

[20] *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.

[21] 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.

[22] 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).

[23] 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.

[24] 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 margin note 59).

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.[25] 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.

[25] 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), whilst the *Political Science Replication Initiative* was launched in *Political Science* (`http://projects.iq.harvard.edu/psreplication/data`).

The Normal Science Bias

THOMAS KUHN FAMOUSLY INSPIRED the introduction of the term paradigm shift to explain that normal science is a very different beast from its non-normal kin (Kuhn, 1962). 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 don'ts. Another prominent double-do in Business and Management is that a study is only taken seriously if (a) the reference list continues over many pages *and* if (b) 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.[26] All references are in the main text and in footnotes, all involving 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*.[27] 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: (1) are the ideas and/or findings presented thought-provoking?; and (2) is the argumentation and/or analysis well done?

[26] 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.

[27] 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, 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.

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 this 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 (1) to discard the double-blind reviewing practice, (2) to thoroughly train reviewers (but see margin note 54), and/or (3) to list the names of the reviewers after publication (see the tentative proposal in the Appendix).

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 multidisciplinary nature (cf. Corley and Gioia, 2011; Pfeffer, 1993).[28]

[28] 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 Manifesto is not the place to engage with this debate.

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.[29] 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 empirical test of their own theory, many discussing the implications for extant theory, many listing managerial implications, *et cetera*. 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).[30]

[29] See, e.g., Birkinshaw et al. (2014); Colquitt and Zapata-Phelan (2007); Davis (2015); Hambrick (2007); Helfat (2007); Tsui (2013).

[30] 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.

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".[31] 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.

[31] 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).

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 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.[32] Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true" (Ioannidis, 2005, 0696).[33]

[32] 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.

[33] 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. After all, 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, 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.[34] 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 pos-

[34] 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 3 in the *Journal of International Business Studies*), of which 57 (87 per cent) do not correct for publication bias at all. The one meta-analysis I have been involved in myself so far (Bogaert et al., 2015) 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.

sible, a research tradition must share a paradigmatic core, such as is the case with organizational ecology (see, e.g., Bogaert et al., 2015). But for those topics within Business and Management that do share such a common core, carrying out meta-analyses would be highly valuable.[35] 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 (Stahl et al., 2009).

[35] From a theoretical angle, Pfeffer and Fong (2005) argue in favor of developing organization theory from "first principles".

Regrettably, the above is easier said than done, for at least two reasons already hinted at above. 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 datasets. 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).

The Market Bias

THAT BUSINESS AND MANAGEMENT, as 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. 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 (™, ®, and ℠). 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.[36] 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? *Et cetera*.

[36] 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; Aguinis et al., 2012; Baum, 2011), 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 (`https://en.wikipedia.org/wiki/San_Francisco_Declaration_on_Research_Assessment`). Cf. Tsui (2013, 2015, 2016).

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 unvoluntary, 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 Multidisciplinary Sciences with an Impact Factor of 3.2 (WoS Journal Citation Reports 2014 Edition)[37] — all quite impressive indeed. Hence, the open access movement does not (yet) really solve the types of problems that are highlighted in this Manifesto's central argumentation.[38]

[37] That is, officially, *Web of Science® Journal Citation Reports®* 2014 Edition.

[38] 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*).

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 (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, 529) rightly observe, "the replicability problems will not be so easily over-come, 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, 647). So, the biases are deeply ingrained in institutionalized cultures, incentives and practices.[39] If there is one stylized fact in Organization Theory, then this is the one: changing such an institutionalized configuration is very hard indeed.[40]

One final issue worth discussing is, given intrinsic human biases, the low-status nature of replication work.[41] In the words of Kane (1984, 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.[42] 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.

[39] This implies the need to seriously rethink 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".

[40] This offers a brilliant opportunity to engage in the usual obligatory referencing to classic studies: DiMaggio and Powell's (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 worldwide academic community or any individual academic institution is close to a mission impossible (cf. margin note 57).

[41] 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").

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

Déjà vu

In the above, only a limited number of references are included, particularly to the classic contributions of Coase (1937), Popper (1959) and Kuhn (1962), to a subset of the large number of warnings regarding the malfunctioning of the scientific community, to a few Business and Management studies, and to many non-scientific sources scraped from the Internet. This is, of course, at odds with any scholarly practice in Business and Management. Not a single reviewer, let alone editor, would accept an essay or Manifesto like this as a serious piece of scholarly work. Hence, in this section, I will do all I can to conform to what the scientific Business and Management community considers to be the bare minimum requirement for any credible output of academic endeavor: extensive, if not exhaustive, referencing. Luckily, this is not a problem at all. Clearly, the above diagnosis is anything but new — quite the contrary. In large and rich disciplines such as Economics, Medicine and Psychology, a similar debate has been ongoing for a very long time (cf. Drotar, 2010).

In Economics, 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, 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, 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 (John et al., 2012). 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).[43] 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*).

[43] 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, 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).

In Medicine, Begg and Berlin (1988, 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., Million and Raoult, 2012; Moreno et al., 2011, 2009; Sterne et al., 2001).

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' rhetorical question (2012: 646–647): "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, 647). In many sciences, independent replication studies represent only one or two per cent 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, 153) report that replications (with or without extensions) "typically constitute less than 10% of published empirical work in the accounting, economics, and finance areas, and 5% 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, 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 per cent in leading Marketing journals (Evanschitzky et al., 2007). Moreover, as Bettis (2012, 110)[44] 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."[45]

[44] See also, e.g., Bedeian et al. (2010); Kacmar (2009); Sutton and Staw (1995); Tsui (2013).

[45] 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.

Is International Business Holier than the Pope?

So, WHAT MAKES MATTERS WORSE, and even outright depressing, is that the battle for scarce publication space can contribute to serious scientific misconduct. Examples of such misconduct are selective outcome reporting, data "massaging", and data fabrication (cf. Moreno et al., 2009),[46] which psychologists refer to as questionable research practices, or QRPs (Pashler and Wagenmakers, 2012), as indicated above (see also Eden, 2010).[47] Of course, the Business and Management community may be the worthiest pupil in the scientific classroom by now, in the 2010s: we, as Business and Management scholars, simply do not do such things anymore. Hence, modern Business and Management is without a serious publication bias. But, given the above arguments, this is rather unlikely. By way of illustration, I briefly discuss the case of International Business, as one of Business and Management's many sub-fields.

[46] Another one is selective sampling (Denrell and Kovács, 2008).

[47] Another striking example of scientific misconduct are fake reviews, or reviews produced by referees closely linked to the authors (or even the authors themselves) using made-up identities and fake email addresses. A recent example, involving the retraction of 64 articles, was revealed by *The Washington Post* in August 2015 (`http://www.washingtonpost.com/news/morning-mix/wp/2015/08/18/outbreak-of-fake-peer-reviews-widens-as-major-publisher-retracts-64-scientific-papers/`).

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*). This is echoed in the field's number-two journal:

"The Journal of World Business is a premier journal in the field of international business [. . .] JWB publishes cutting-edge research that reflects important developments in the global business environment and advances new theoretical directions [. . .] The journal encourages submissions that break new ground or demonstrate novel or counterintuitive findings in relation to established theories or assumptions" (Webpage of the *Journal of World Business*). And a critical statement on *Management International Review*'s aims and scope page signals that "Editors are especially interested in manuscripts that break new ground rather than papers that only make incremental contributions."

The first vehicle to correct for publication biases are systematic meta-analyses. The four meta-analyses published in the *Journal of International Business Studies*, to date (1970–now), prove that this can be done in the International Business domain (Meyer and Sinasi, 2009; Peterson and Jolibert, 1995; Stahl et al., 2009; Tihanyi et al., 2005). However, the extremely low number of four is very disappointing indeed.[48] For one reason or another, meta-analyses in International Business are far and between. The second vehicle are replication studies. An electronic search in the *Journal of International Business Studies*' e-archive (1970–now) does not give a single hit. 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.[49] 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 per cent of the total. The number of nulls is 13 (9 per cent), and the number of negatives is 5 (4 per cent). Rather strikingly, in 20 of these 28 hypotheses-testing studies (71 per cent), *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.

The outcome of this crude counting exercise is discouraging. It appears that also International Business is not exempt from the very

[48] A well-published Business and Management scholar recently had a meta-analysis, after a revision round, rejected by a top Business and Management journal's editor with the argument that the journal does not really like (to publish) meta-analyses — *quod non* (personal email exchange). Another colleague struggles to pass the reviewing process at another top Business and Management outlet as the editor requires that the meta-analysis contributes to new theory — another *quod non* (personal email exchange).

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

forceful anti-Popperian development. Nulls and negatives are a rarity, meta-analyses are far and between, and independent replication studies are nowhere to be seen. Understandably so, all International Business scholars seem to chase for new discoveries — for "cutting-edge" and "ground-breaking" positives. We all aspire to be a scientific Messi, albeit only a few of us can hope to achieve that status. In all likelihood, the state of affairs in International Business (and in the other sub-disciplines of Business and Management) is even worse than that in Economics, Psychology, or (Life and Physical) Sciences, given Fanelli's (2010) finding that the addiction to positives is increasing in disciplines in lower positions on the scientific status ladder. Moreover, as (International) Business and Management is dominated by small sample sizes, small effect sizes and isolated pieces of research, rather than large sample sizes, large effect sizes and grand collaborative projects, the likelihood of reported false positives can possibly be higher than that in Medicine, and the Life and Physical Sciences (Ioannidis, 2005).[50] 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, Eden (2002, 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.

[50] Ioannidis (2005) identifies two further features of a scientific field that increase the likelihood of finding false positives: a large number and low selection of tested relationships, and great flexibility in designs, definitions, outcomes, and analytical models. On both accounts, this does not bode well for Business and Management.

[51] 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.

[52] See, e.g., Evanschitzky et al. (2007); Mezias and Regnier (2007) (in *Strategic Organization* — déjà vu indeed); Renkewitz et al. (2011); Schooler (2011)

How to Escape from this Deadlock?

So, WHAT CAN WE, AS A COLLECTIVE of Business and Management scholarship, do to escape from this dismal state of affairs?[51] By and large, similar measures are suggested across a wide range of disciplines. In this Manifesto, I would like to briefly discuss seven key measures.[52].

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*.[53] Essential is that, in so doing, scholars make

[53] 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 (`https://www.socialscienceregistry.org/`), which is mainly used for field experiments (and not for non-experimental field work).

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. [54]

4. Fourth, all raw data, protocols and data analysis codes, or any other relevant material (for instance, for computer simulation studies, this should involve software programming codes), 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.[55] 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.

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, 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.[56]

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.

[54] 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; however, this is anything but easy (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).

[55] 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).

[56] 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 beta 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, for example, the recent book by Imbens and Rubin (2015)). For instance, given the often small sample sizes in Business and Management, one would expect tests to be validated using bootstrapping methods, so one does not rely on asymptotic results.

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. 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 Business and Management. 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, 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."[57]

[57] 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 Business and Management (cf., in Economics, Dewald et al. (1986); McCullough et al. (2006, 2008)). We as 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).

The Way Forward

Above, I suggested that the time is right to revitalize Business and Management as a scientific community by establishing three new journals: the *Journal of Business and Management Replication Studies*, the *Journal of Heterodox Perspectives in Business and Management*, and the *Journal of Business and Management Meta-Analyses*. Launching these three new journals is not very likely to happen in the near future if we, as a community, fail to organize collective action.[58] Alternatively, any of the current outlets that takes the above diagnosis seriously, could add three new sections to the journal: (a) Meta-Analyses Section; (b) Replication Studies Section; and (c) Heterodox Perspectives Section. This would imply an attempt to change current practices from within, building on the legitimacy of existing journals. Would the outlet's publication space be too limited to accommodate all this, the journal could follow in the footsteps of PLoS and launch a separate open access e-journal for this — say, titled *Business and Management Letters* — or three separate open access e-journals dedicated to implement the above. Recent initiatives like the *Academy of Management Discoveries* and the *Journal of Business Ventures Insights* prove that this can be done.

[58] Would a well-reputed association decide to invest in launching this set of e-journals, like the European Association of Social Psychology did by establishing the *Comprehensive Results in Social Psychology* outlet, this could create momentum by mobilizing the immediate backing of an institutionalized community.

In any case, all journals should try to replace bad by good scientific practices, apart from providing a platform for meta-analyses, nulls and negatives, failed and successful replication studies, and heterodox perspectives. A few tailor-made initiatives could be very powerful. For one, authors of accepted studies must put their raw data, protocols and data analysis codes online in a tailor-made journal replication material archive (perhaps, in collaboration with an archive consortium). On top of that, journals can decide to hire a small staff to annually run a few replication analyses, the results of which are reported in the journal's new year's issue.[59] This little staff is headed by a dedicated Replication Section Editor. Hopefully, with one sheep daring to pass the dyke (once a Dutchman, always a Dutchman), there will be many more following this example, including the field's top outlets. Even more, new ways of publishing might be developed, different from the current WoS Impact Factor journals' straightjackets. In the Appendix, I include a short and tentative template of what the key elements of a new mode of publishing might look like, coined Scientific Wikipedia, for now.

[59] Say, roughly five per cent of the published studies could be replicated, using the original raw data. In all likelihood, this practice will have a preventive effect, as scholars knowing that this may happen will think twice before engaging in questionable research practices. See also Fox's (1994) idea to implement random data audits (cf. Bentler, 2007).

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: `https://www.change.org`. This petition does not speak to the Business and Management community alone — quite the contrary. After all, the issues discussed in this Manifesto 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. Evanschitzky et al., 2007; Hubbard and Vetter, 1996). Hence, all those who would support such a set of initiatives, inside and outside the 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.

It is time for radical change.

APPENDIX: Scientific Wikipedia

A Brief and Tentative Template

Wikipedia is a free encyclopedia built collaboratively using wiki software. Similarly, Scientific Wikipedia (SW)[60] could be developed dynamically and collaboratively by the scholarly community. After an initial screen by an editorial board, any submission that passes a minimum threshold of scientific rigor is posted on SW without any immediate need for changes. Subsequently, reviewers are asked to write and post non-anonymous comments. Moreover, all SW-reading scholars are invited to post non-anonymous comments.[61] At any time, the authors may decide to upload a revised version of their original paper, to withdraw the original study, or to write a separate response. To each accepted submission, a dynamic account is attached, providing a series of statistics (number of times cited and downloaded, number of revisions, links to papers citing this submission, *et cetera*), similar to that offered by Research Gate (`http://www.researchgate.net`). Of course, SW would adopt all the practices referred to in the main text (e.g., separate sections for meta-analyses and replication studies, a pre-registration repository, a dedicated replication team, and material upload requirements). SW will be free, not charging any fees, and strict open access policies are pursued.[62]

[60] This is just a working name. Suggestions for a better one are more than welcome.

[61] In a way, this implies a return to classic practices. Academic journals were established to facilitate open scientific dialogue, replacing the exchange of bilateral (and non-anonymous) letters among scholars.

[62] Would, say, ten business schools be willing to invest an annual sum of €50,000 in SW, a healthy budget will be available to run this initiative.

References

Adler, N. and A. Harzing: 2009, 'When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings'. *Academy of Management Learning & Education* **8**, 72–95.

Aguinis, H., I. Suarez-Gonzalez, G. Lannelongue, and H. Joo: 2012, 'Scholarly Impact Revisited'. *Academy of Management Perspectives* **26**, 105–132.

Barker, V. and M. Mone: 1994, 'Retrenchment: Cause of Turnaround or Consequence of Decline?'. *Strategic Management Journal* **15**, 395–405.

Baum, J.: 2011, 'Free-Riding on Power Laws: Questioning the Validity of the Impact Factor as a Measure of Research Quality in Organization Studies'. *Organization* **18**, 449–466.

Bedeian, A.: 2003, 'The Manuscript Review Process: The Proper Roles of Authors, Referees, and Editors'. *Journal of Management Inquiry* **12**, 331–338.

Bedeian, A., S. Taylor, and A. Miller: 2010, 'Management Science on the Credibility Bubble: Cardinal Sins and Various Misdemeanors'. *Academy of Management Learning & Education* **9**, 715–725.

Begg, C. and J. Berlin: 1988, 'Publication Bias: A problem in Interpreting Medical Data'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **151**, 419–463.

Bentler, P.: 2007, 'On Tests and Indices for Evaluating Structural Models'. *Personality and Individual Differences* **42**, 825–829.

Bettis, R.: 2012, 'The Search for Asterisks: Comprised Statistical Tests and Flawed Theories'. *Strategic Management Journal* **33**, 108–113.

Birkinshaw, J., M. Healey, R. Suddaby, and K. Weber: 2014, 'Debating the Future of Management Research'. *Journal of Management Studies* **51**, 38–55.

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

Bogaert, S., C. Boone, G. Negro, and A. van Witteloostuijn: 2015, 'Organizational Form Emergence: A Meta-Analysis of the Ecological Theory of Legitimation'. *Journal of Management*. (forthcoming).

Bradford De Long, J. and K. Lang: 1992, 'Are all Economic Hypotheses False?'. *Journal of Political Economy* **100**, 1257–1272.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg: 2012, 'Star Wars: The Empirics Strike Back'. Working Paper 2012-29, Paris School of Management, Paris.

Callaham, M. and C. McCulloch: 2011, 'Longitudinal Trends in the Performance of Scientific Peer Reviewers'. *Annals of Emergency Medicine* **57**, 141–148.

Callaham, M. and J. Tercier: 2007, 'The Relationship of Previous Training and Experience of Journal Peer Reviewers to Subsequent Review Quality'. *PLoS Medicine* **4**, 0032–0040.

Card, D. and A. Krueger: 1995, 'Time-Series Minimum Wage Studies: A Meta-Analysis'. *American Economic Review Papers and Proceedings* **85**, 238–243.

Coase, R.: 1937, 'The Nature of the Firm'. *Economica* **16**, 386–405.

Colquitt, J. and C. Zapata-Phelan: 2007, 'Trends in Theory Building and Theory Testing: A Five-Decade Study of the *Academy of Management Journal*'. *Academy of Management Journal* **50**, 1281–1303.

Corley, K. and D. Gioia: 2011, 'Building Theory about Theory Building: What Constitutes a Theoretical Contribution?'. *Academy of Management Review* **36**, 12–32.

Davis, G.: 2015, 'Editorial Essay: What is Organizational Research for?'. *Administrative Science Quarterly* **60**, 179–188.

Delbridge, R. and P. Fiss: 2013, 'Editors' Comments: Styles of Theorizing and the Social Organization of Knowledge'. *Academy of Management Review* **38**, 325–331.

Delmar, F.: 2015, 'A Response to Honig and Samuelsson (2014)'. *Journal of Business Venturing Insights* **3**, 1–4.

Denrell, J. and B. Kovács: 2008, 'Selective Sampling of Empirical Settings in Organizational Studies'. *Administrative Science Quarterly* **53**, 109–144.

Dewald, W., J. Thursby, and R. Anderson: 1986, 'Replication in Empirical Economics: The *Journal of Money, Credit and Banking* Project'. *American Economic Review* **76**, 587–603.

Dikova, D. and A. van Witteloostuijn: 2007, 'Acquisition versus Greenfield Foreign Entry: Diversification Mode Choice in Central and Eastern Europe'. *Journal of International Business Studies* **38**, 1013–1033.

DiMaggio, P. and W. Powell: 1983, 'The Iron Cage Revisited: Collective Rationality and Institutional Isomorphism in Organizational Fields'. *American Sociological Review* **48**, 147–160.

Doucouliagos, H. and T. Stanley: 2009, 'Publication Bias in Minimum-Wage Research? A Meta-Regression Analysis'. *British Journal of Industrial Relations* **47**, 406–428.

Drotar, D.: 2010, 'Editorial: A Call for Replications of Research in Pediatric Psychology and Guidance for Authors'. *Journal of Pediatric Psychology* **35**, 801–805.

Duval, S. and R. Tweedie: 2000, 'A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis'. *Journal of American Statistical Association* **95**, 89–98.

Eden, D.: 2002, 'Replication, Meta-Analysis, Scientific Progress, and *AMJ*'s Publication Policy'. *Academy of Management Journal* **45**, 841–846.

Eden, L.: 2010, 'Letter from the Editor-in-Chief: Scientists Behaving Badly'. *Journal of International Business Studies* **41**, 561–566.

Evanschitzky, H., C. Baumgarth, R. Hubbard, and J. Armstrong: 2007, 'Replication Research's Disturbing Trend'. *Journal of Business Research* **60**, 411–415.

Fanelli, D.: 2010, '"Positive" Results Increase Down the Hierarchy of the Sciences'. *PLoS ONE* **5**, e10068.

Ferguson, C. and M. Heene: 2012, 'A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null'. *Perspectives on Psychological Science* **7**, 555–561.

Fox, M.: 1994, 'Scientific Misconduct and Editorial and Peer Review Processes'. *Journal of Higher Education* **65**, 298–309.

Geyskens, I., R. Krishnan, J. Steenkamp, and P. Cunha: 2009, 'A Review and Evaluation of Meta-Analysis Practices in Management Research'. *Journal of Management* **35**, 393–419.

Ghoshal, S.: 2005, 'Bad Management Theories Are Destroying Good Management Practices'. *Academy of Management Learning & Education* **4**, 75–91.

Golden, M.: 1995, 'Replication and Non-Quantitative Research'. *PS: Political Science and Politics* **28**, 481–483.

Grey, C.: 2010, 'Organizing Studies: Publications, Politics and Polemic'. *Organization Studies* **31**, 677–694.

Hambrick, D.: 2007, 'The Field of Management's Devotion to Theory: Too Much of a Good Thing?'. *Academy of Management Journal* **50**, 1346–1352.

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

Hannan, M. and J. Freeman: 1977, 'The Population Ecology of Organizations'. *American Journal of Sociology* **82**, 929–964.

Hannan, M. and J. Freeman: 1984, 'Structural Inertia and Organizational Change'. *American Sociological Review* **49**, 149–164.

Head, M., L. Holman, R. Lanfear, A. Kahn, and M. Jennions: 2015, 'The Extent and Consequences of P-Hacking in Science'. *PLoS Biology* **13**, e10022106.

Helfat, C.: 2007, 'Stylized Facts, Empirical Research and Theory Development in Management'. *Strategic Organization* **5**, 185–192.

Hoenig, J. and D. Heisey: 2001, 'The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis'. *The American Statistician* **55**, 1–6.

Honig, B. and M. Samuelsson: 2014, 'Data Replication and Extension: A Study of Business Planning and Venture-Level Performance'. *Journal of Business Venturing Insights* **1–2**, 18–25.

Hubbard, R. and J. Armstrong: 1997, 'Publication Bias against Null Results'. *Psychological Reports* **80**, 337–338.

Hubbard, R. and D. Vetter: 1996, 'An Empirical Comparison of Published Replication Research in Accounting, Economics, Finance, Management, and Marketing'. *Journal of Business Research* **35**, 153–164.

Hubbard, R., D. Vetter, and E. Little: 1998, 'Replication in Strategic Management: Scientific Testing for Validity, Generalizability, and Usefulness'. *Strategic Management Journal* **19**, 243–254.

Imbens, G. and D. Rubin: 2015, *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.

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

Ioannidis, J.: 2005, 'Why Most Published Research Findings Are False'. *PLoS Medicine* **2**, e124.

Ioannidis, J.: 2012, 'Why Science is Not Necessarily Self-Correcting'. *Perspectives on Psychological Science* **7**, 645–654.

Isaeva, N., R. Bachmann, A. Bristow, and M. Saunders: 2015, 'Why the Epistemologies of Trust Researchers Matter'. *Journal of Trust Research*. (forthcoming).

Jefferson, T., P. Alderson, E. Wagner, and F. Davidoff: 2002, 'Effects of Editorial Peer Review: A Systematic Review'. *Journal of the American Medical Association* **287**, 1–4.

John, L., G. Loewenstein, and D. Prelec: 2012, 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth-Telling'. *Psychological Science* **23**, 524–532.

Jonas, K. and J. Cesario: 2015, 'How Can Preregistration Contribute to Research in Our Field?'. *Comprehensive Results in Social Psychology*.

Kacmar, K.: 2009, 'From the Editors: An Ethical Quiz'. *Academy of Management Journal* **52**, 432–434.

Kahneman, D.: 2011, *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.

Kane, E.: 1984, 'Why Journal Editors Should Encourage the Replication of Applied Econometrics Research'. *Quarterly Journal of Business & Economics* **23**, 3–8.

Kerr, N.: 1998, 'HARKing: Hypothesizing After the Results are Known'. *Personality and Social Psychology* **2**, 196–217.

Kuhn, T.: 1962, *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

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

Levy, O., S. Taylor, N. Boyacigiller, T. Bodner, M. Peiperl, and S. Beechler: 2015, 'Perceived Senior Leadership Opportunities in MNCs: The Effect of Social Hierarchy and Capital'. *Journal of International Business Studies* **46**, 285–307.

Lexchin, J., L. Bero, B. Djulbegovic, and O. Clark: 2003, 'Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review'. *British Medical Journal* **326**, 1167–1170.

Lindebaum, D.: 2015, 'Critical Essay: Building New Management Theories on Sound Data? The Case of Neuroscience'. *Human Relations*. (forthcoming).

van Maanen, J.: 1989, 'Some Notes on the Importance of Writing in Organization Studies'. *Harvard School Research Colloquium* **27**, 133–143.

Makel, M., J. Plucker, and B. Hegarty: 2012, 'Replications in Psychology Research: How Often do They Really Occur?'. *Perspectives on Psychological Science* **7**, 537–542.

Mayer, M. and R. Whittington: 2003, 'Diversification in Context: A Cross-National and Cross-Temporal Extension'. *Strategic Management Journal* **24**, 773–781.

McCullough, B., K. McGeary, and T. Harrison: 2006, 'Lessons from the *JMCB* Archive'. *Journal of Money, Credit and Banking* **38**, 1093–1107.

McCullough, B., K. McGeary, and T. Harrison: 2008, 'Do Economics Journal Archives Promote Replicable Research?'. *Canadian Journal of Economics* **41**, 1406–1420.

Melander, H., J. Ahlqvist-Rastad, G. Meijer, and B. Beermann: 2003, 'Evidence B(i)ased Medicine — Selective Reporting from Studies Sponsored by Pharmaceutical Industry: Review of Studies in New Drug Applications'. *British Medical Journal* **326**, 1171–1173.

Meyer, K. and E. Sinasi: 2009, 'When and Where Does Foreign Direct Investment Generate Positive Spillovers? A Meta-Analysis'. *Journal of International Business Studies* **40**, 1075–1094.

Mezias, S. and M. Regnier: 2007, 'Walking the Walk as well as Talking the Talk: Replication and the Normal Science Paradigm in Strategic Management Research'. *Strategic Organization* **5**, 283–296.

Miller, D.: 2007, 'Paradigm Prisons, or in Praise of Atheoretic Research'. *Strategic Organization* **5**, 177–184.

Million, M. and D. Raoult: 2012, 'Publication Biases in Probiotics'. *European Journal of Epidemiology* **27**, 885–886.

Mingers, J. and H. Willmott: 2013, 'Taylorizing Business School Research: On the "One Best Way" Performative Effects of Journal Ranking Lists'. *Human Relations* **66**, 1051–1073.

Mittelstaedt, R. and T. Zorn: 1984, 'Economic Replications: Lessons from the Experimental Sciences'. *Quarterly Journal of Business and Economics* **23**, 9–15.

Moreno, S., A. Sutton, A. Ades, N. Cooper, and K. Abrams: 2011, 'Adjusting for Publication Biases Across Similar Interventions Performed Well When Compared with Gold Standard Data'. *Journal of Clinical Epidemiology* **64**, 1230–1241.

Moreno, S., A. Sutton, E. Turner, K. Abrams, N. Cooper, T. Palmer, and A. Ades: 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 Online First* **339**, b2981.

Neuroskeptic: 2012, 'The Nine Circles of Scientific Hell'. *Perspectives on Psychological Science* **7**, 643–644.

Nuijten, M., M. van Assen, C. Veldkamp, and J. Wicherts: 2015, 'The Replication Paradox: Combining Studies can Decrease Accuracy of Effect Size Estimates'. *Review of General Psychology*. (forthcoming).

Nuzzo, R.: 2014, 'Statistical Errors: *p*-Values, the 'Gold Standard' of Statistical Validity, Are Not as Reliable as Many Scientists Assume'. *Nature* **506**, 150–152. http://www.nature.com/news.

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

Open Science Collaboration: 2015, 'Estimating the Reproducibility of Psychological Science'. *Science* **349**, aac4716.

Parker, M. and R. Thomas: 2011, 'What Is a Critical Journal?'. *Organization* **18**, 419–427.

Pashler, H. and E. Wagenmakers: 2012, 'Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?'. *Perspectives on Psychological Science* **7**, 528–530.

Peterson, R. and A. Jolibert: 1995, 'A Meta-Analysis of Country-of-Origin Effects'. *Journal of International Business Studies* **26**, 883–900.

Pfeffer, J.: 1993, 'Barriers to the Advance of Organizational Science: Paradigm Development as a Dependent Variable'. *Academy of Management Review* **18**, 599–620.

Pfeffer, J. and C. Fong: 2005, 'Building Organization Theory from First Principles'. *Organization Science* **16**, 372–388.

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

Renkewitz, F., H. Fuchs, and S. Fiedler: 2011, 'Is There Evidence of Publication Bias in JDM Research?'. *Judgment and Decision Making* **6**, 870–881.

Rosenthal, R.: 1979, 'An Introduction to the File Drawer Problem'. *Psychological Bulletin* **86**, 638–641.

Sarasvathy, S.: 2003, 'Entrepreneurship as a Science of the Artificial'. *Journal of Economic Psychology* **24**, 203–220.

Schooler, J.: 2011, 'Unpublished Results Hide the Decline Effect'. *Nature* **470**, 437.

Schroter, S., S. Black, F. Evans, L. Godlee, L. Osorio, and R. Smith: 2008, 'What Errors Do Peer Reviewers Detect, and Does Training Improve their Ability to Detect Them?'. *Journal of the Royal Society of Medicine* **101**, 507–514.

Schultz, M. and M. Hatch: 1996, 'Living with Multiple Paradigms: The Case of Paradigm Interplay in Organizational Culture Studies'. *Academy of Management Review* **21**, 529–557.

Shapiro, D., R. Kirkman, and H. Courtney: 2007, 'Perceived Causes and Solution of the Translation Problem in Management Research'. *Academy of Management Journal* **50**, 249–266.

Shleifer, A.: 2012, 'Psychologists at the Gate: A Review of Daniel Kahneman's *Thinking, Fast and Slow*'. *Journal of Economic Literature* **50**, 1080–1091.

Simmons, J., L. Nelson, and U. Simonsohn: 2011, 'False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant'. *Psychological Science* **22**, 1356–1366.

Sorge, A. and K. Rothe: 2011, 'Resource Dependence and Construction, and Macro and Micro Politics in Transnational Enterprises and Alliances: The Case of Jet Engine Manufacturers in Germany'. In: M. Geppert and C. Dörrenbächer (eds.): *Politics and Power in the Multinational Corporation: The Role of Institutions, Interests and Identities*. Cambridge University Press, pp. 41–71.

Stahl, G., M. Maznevski, A. Voigt, and K. Jonsen: 2009, 'Unraveling the Effects of Cultural Diversity in Teams: A Meta-Analysis of Research on Multicultural Work Groups'. *Journal of International Business Studies* **40**, 690–709.

Stanley, T.: 2001, 'Wheat from Chaff: Meta-Analysis as Quantitative Literature Review'. *Journal of Economic Perspectives* **15**, 131–150.

Stanley, T.: 2008, 'Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection'. *Oxford Bulletin of Economics and Statistics* **70**, 103–127.

Starbuck, W.: 2003, 'Turning Lemons into Lemonade: Where is the Value in Peer Reviews?'. *Journal of Management Inquiry* **12**, 344–351.

Starbuck, W.: 2004, 'Why I Stopped Trying to Understand the Real World'. *Organization Studies* **25**, 1233–1254.

Stengers, I.: 2000, *The Invention of Modern Science*, Vol. 19. Minneapolis, MN: University of Minnesota Press.

Sterling, T.: 1959, 'Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance'. *Journal of the American Statistical Association* **54**, 30–34.

Sterne, J., M. Egger, and G. Smith: 2001, 'Systematic Reviews in Health Care: Investigating and Dealing with Publication and other Biases in Meta-Analysis'. *British Medical Journal* **323**, 101–105.

Sutton, R. and B. Staw: 1995, 'What Theory Is Not'. *Administrative Science Quarterly* **40**, 371–384.

The Economist: 2013, 'Trouble at the Lab'. `http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble`.

Tihanyi, L., D. Griffith, and C. Russell: 2005, 'The Effect of Cultural Distance on Entry Mode Choice, International Diversification, and MNE Performance: A Meta-Analysis'. *Journal of International Business Studies* **36**, 270–283.

Tsang, E. and K. Kwan: 1999, 'Replication and Theory Development in Organizational Science: A Critical Realist Perspective'. *Academy of Management Review* **24**, 759–780.

Tsui, A.: 2013, 'On Compassion in Scholarship: Why Should We Care?'. *Academy of Management Review* **38**, 167–181.

Tsui, A.: 2015, 'Reconnecting with the Business World: Socially Responsible Scholarship'. *EFMD Global Focus* **9**, 36–39.

Tsui, A.: 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).* (forthcoming).

Tsui, A. and J. Hollenbeck: 2009, 'Successful Authors and Effective Reviewers Balancing Supply and Demand in the Organizational Sciences'. *Organizational Research Methods* **12**, 259–275.

van de Ven, A., S. Ang, A. Arino, P. Bamberger, C. LeBaron, C. Miller, and F. Milliken: 2015, 'Welcome to the *Academy of Management Discoveries (AMD)*'. *Academy of Management Discoveries* **1**, 1–4.

Walsh, J., K. Weber, and J. Margolis: 2003, 'Social Issues and Management: Our Lost Cause Found'. *Journal of Management* **29**, 859–881.

Wicherts, J., M. Bakker, and D. Molenaar: 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* **6**, e26828.

Wigboldus, D. and R. Dotsch: 2015, 'Encouraging Playing with Data and Discourage Questionable Reporting Practices'. *Psychometrika*. (forthcoming).

van Witteloostuijn, A.: 2015, 'Toward Experimental International Business: Unraveling Fundamental Causal Linkages'. *Cross-Cultural and Strategic Management (*formerly known as *Cross Cultural Management)*. (forthcoming).

van Witteloostuijn, A.: 2016, 'Publishing Null and Negative Findings: Moving Away from Biased Publication Practices'. *Cross-Cultural and Strategic Management (*formerly known as *Cross Cultural Management)*. (forthcoming).

Yong, E.: 2013, 'Psychologists Strike a Blow for Reproducibility: Thirty-Six Labs Collaborate to Check 13 Earlier Findings'. *Nature*.

Zimmerman, C.: 2015, 'On the Need for a Replication Journal'. Working Paper 2015-016A, Federal Reserve Bank of St. Louis, St. Louis, MO.