



Null Findings, Replications and Preregistered Studies in Business Ethics Research

Julia Roloff¹ · Michael J. Zyphur²

Received: 5 May 2017 / Accepted: 26 March 2018 / Published online: 5 April 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Comment from the Editors in Chief, Journal of Business Ethics

In this essay, editors at the Journal of Business Ethics, Julia Roloff and Mike Zyphur, explore the practice of preregistered research (i.e. wherein research plans are assessed before data collection starts) and propose the trial of a preregistration procedure at the journal. Together with several other Journal of Business Ethics editors, they will edit a special issue designed to experiment with their suggested protocol and build our knowledge and capacity around preregistration. We hope to get a number of papers that actually use the preregistered protocol being trialled, as well as a number of papers that critically examine the idea of preregistration, publication of null results and all of the ethical issues associated with these ideas.

It is both timely and appropriate for the journal to be pursuing the topic of preregistered studies. The Journal of Business Ethics is marked by its breadth and depth of scholarship and expertness of editors across many fields. This gives us the opportunity to be adventurous, to explore new ideas and take some risks with the new practices. We would like to explicitly encourage quantitative research that is both focused on ethical issues and undertaken in an ethical manner. We are not taking the position that preregistered reporting of studies is the only way to achieve this goal, nor that it necessarily will achieve this goal. However, we would like

to explore these possibilities. In short, we offer these ideas as an experiment. We are committed to making Journal of Business Ethics a place where innovation and experimentation are possible. We are grateful to the authors of this essay and a number of our editors for their important contributions to this effort.

Introduction

Are academics conducting research to create practical knowledge or merely to publish? Does a focus on ‘getting published’—where it is present—improve or limit choices of research topics, samples and methodologies? Are reviewers and editors appreciating the results of studies based on their practical implications or do they favour papers that offer support for supposedly a priori hypotheses? Are prevailing academic norms facilitating best practices in research and produce trustworthy and meaningful outcomes? These questions have been taken up in many recent articles, editorials and academic initiatives across the social and physical sciences, often with the troubling answer that an interest in publishing, combined with review and editorial practices that favour supported hypotheses, leads to questionable research practices that are widespread and reduce the trustworthiness of research (see Banks et al. 2016; Bettis 2012; Community for Responsible Research in Business and Management 2017).

How did this happen? There are several factors contributing to the problem. Journal editors and reviewers ask for honesty but incentivize dishonesty by treating a study as relevant when hypotheses are supported rather than whether the study as a whole is practically useful. In turn, researchers correctly anticipate a publication bias wherein editors and reviewers prefer studies in which hypotheses are supported by the data (Byington and Felps 2017). Worried about job security, academics then feel the pressure to engage in questionable research practices such as foraging for statistically

✉ Julia Roloff
julia.roloff@rennes-sb.com

Michael J. Zyphur
mzyphur@unimelb.edu.au

¹ Rennes School of Business, Centre for Responsible Business, 2, Rue Robert d’Abrissel, CS 76522, 35065 Rennes, France

² Department of Management and Marketing, Faculty of Business and Economics, University of Melbourne, Melbourne, Australia

significant effects and reporting unexpected findings as if they were expected, or simply deciding not to write up studies that have ‘unpublishable’ results (Banks et al. 2016). The result is that researchers often censor results and/or present their findings in bad faith, as if something unexpected was expected all along.

Ultimately, such practices undermine trust in published research as readers of academic journals cannot be confident in whether published articles are the product of research conducted and described in good faith. This is to say that published research may be untrustworthy because it has not been conducted and described on the shared terms of a scientific community, terms that are meant to enable honesty and transparency in research practices. Consequently, managers, policy-makers and other stakeholders may hesitate to put scientific conclusions to practical use. On a personal basis, the problem of living an honest life confronts researchers, as they are motivated to see research as a publishing ‘game’ rather than as a collective endeavour that has at its core an ethical responsibility to be civically useful—this is a severe disappointment for many Ph.D. students who are confronted by this careerist ‘game’ (Bettis 2012).

A passionate discussion has emerged in business research regarding how this problem can and should be solved. For example, contributors to a special issue in the *Academy of Management Learning and Education* (Bergh et al. 2017; Byington and Felps 2017; Schwab and Starbuck 2017), a group of guest editors at the *Journal of Management* (Banks et al. 2016) and a former editor of *Administrative Science Quarterly* (Starbuck 2016) outline what journal editors, publishing houses, individual researchers and methodology teachers can do to: (1) foster honest and transparent research practices on the shared terms of a scientific community; (2) detect questionable practices; and (3) enable replication studies and the publication of null findings—all of which may improve trust in research.

The *Journal of Business Ethics* (JBE) seeks to contribute to this discussion and address the challenge of trust in research. In this editorial, we explore the possibility of introducing new procedures at JBE to address this problem and to reaffirm existing commitments such as being open to publishing null findings and replication studies. Specifically, we discuss the potential of preregistered research protocols wherein research plans are reviewed before data collection starts. This procedure is meant to separate evaluations of study quality from those of its results. Moreover, JBE commits to testing this protocol by announcing a Special Issue with this focus. The Special Issue will welcome empirical contributions following this protocol as well as articles that offer critical inquiry into this practice and the many issues that constellate with trust and transparency in science and its relationship with society (e.g. Jasanoff 2004; Poovey 1998; Porter 1986; Strathern 2000; Tsoukas 1997; see also a recent

special issue of *Science, Technology, and Human Values*: Leonelli et al. 2017).

In the following section, we discuss some of the processes and practices that contribute to a crisis of trust in research. Afterwards, we introduce a protocol for preregistered research that encourages more transparent and honest research practices. Moreover, we announce a call for contribution to the Special Issue which tests and critically discusses this editorial protocol.

A Crisis of Trust in Research

Trust in researchers and the usefulness of their results is central to ensuring the quality and relevance of research. Although there are sources of intrinsic uncertainty that will always thwart the goal of singularly ‘true’ theories or perfect predictions, researchers can gather evidence and make inferences that offer more or less warrant for specific assertions and the practical activities with which they may be associated (van Fraassen 2008), such as particular ways of organizing, making decisions, or more generally living an organizational life (du Gay 2015). One problem here is that questionable research practices—especially associated with the dishonest reporting of results—are extremely common and some observers fear that a large portion of the findings published in management journals are misleading, if not outright wrong on the terms of the very epistemic logics they use for their own legitimacy (e.g. based on the logic of hypothesis testing with statistic and probability; Banks et al. 2016; Schwab and Starbuck 2017). In turn, questions about honesty and trust in research have become increasingly important (see Fanelli 2011).

The fact that honesty is central to science within and across communities is well known, as trust provides a basis for the creation and dissemination of scientific practices and their associated discourses (see Jasanoff 2004, 2009, 2010, 2014; Poovey 1998; Porter 1996; Shapin and Schaffer 1985). Yet, how trust *should* be mapped to the practice of research in its current institutionalized form has been less well investigated—this is an open ethical question. Many normative initiatives related to the proper conduct and communication of research focus on ethical commitments that masquerade as epistemic motivations towards validity or a singular, abstracted ‘truth’ vis-à-vis accurately testing the correspondence among theories and an external reality (e.g. ‘valid’ measures or hypothesis tests as judged by whether they reduce ‘errors in inference’; Cortina and Landis 2011; Cumming 2014; Morey et al. 2014). This focus on epistemology can be useful if it is contextualized in a broader milieu of different epistemic cultures that embody different approaches to research (e.g. Knorr-Cetina 2009, 2013) and connected to the different practical implications of specific

epistemic commitments (see Dewey 1920, 1938; Fish 1985, 2003; van Fraassen 2008).

A common way to understand dilemmas associated with trust and questionable research practices is through the logic of statistics and probability as these relate to hypothesis testing—especially via notions of errors in inference (e.g. Type I/II errors). Although we periodically use these terms to connect with the epistemic communities that appreciate logics of inference under a probabilistic uncertainty, we do not unconditionally endorse this understanding of the basis for trusting research results. As we note in what follows, no existing work offers good reasons why probabilistic uncertainty might have a clear ethical and practical relationship with questionable research practices and their outcomes *without tautologically relying on a logic of probabilistic uncertainty itself*. As such, in our call for contributions, we note JBE's openness to critical evaluations of trust in research and how this relates to typical quantitative logics and practices, as well as how these logics and practices obscure the problem of acting and speaking in good faith, including by engaging in speech acts that are meant to serve as testimony and/or promises (i.e. reporting research results). In brief, communities are built on trust, as embedded in discourse, technologies and shared practices. Scientific communities are no different. Speaking and acting in good faith are essential for being able to practically *act* collectively as researchers, perhaps especially at a journal like JBE that is devoted to ethics.

To this end, it is notable that in discussions about researchers' integrity and questionable research practices, honesty and trust are often treated peripherally, as if the ethical commitments of a particular epistemology come first and then trust and its attendant benefits follow as communities 'discover' the 'truth' of an epistemic orientation. This puts the cart before the horse. As history shows, trust is a foundational pillar upon which epistemic logics and their cultures can be built and then accepted and institutionalized by a society (Poovey 1998; Porter 1996; Shapin and Schaffer 1985; Shapin 1994, 2009). Therefore, trust within and across communities and their institutions should be the focus of initiatives meant to map scientific practices onto journal submission and reviewing protocols. To grapple with this, we treat three reasons why trust in published research may be hindered: questionable research practices, publication bias and bias against evidence questioning a theory's validity. JBE's openness to publishing null findings, replications and preregistration is meant to tackle these issues.

Questionable Research Practices

Across five empirical studies, Banks et al. (2016) investigated how often management researchers engage in questionable research practices. They found that the reporting

of the analysis process and research results lacked transparency—whereas the falsification of data appeared to be very rare (Banks et al. 2016, supplement DS7), with only three of 749 active management researchers (0.4%) admitting to falsifying data (although those capable of falsifying data may have lied in response to the questionnaire; John et al. 2012). In the same study, 11.1% of management researchers reported that they had rounded off a p value (e.g. from 0.054) to appear to be at or below the $p = 0.05$ threshold; 28.5% excluded data (such as outliers) after looking at the impact of doing so on the results; 33.3% included or excluded control variables based on the statistical significance of the results; 49.6% admitted to HARKing, or the presentation of post hoc hypotheses as a priori hypotheses; and 49.7% selectively reported hypotheses based on whether or not they were statistically significant. Furthermore, these latter values may be as high as 90% when inducements for truth telling are implemented (see John et al. 2012). Banks et al. (2016) also reported that about a third of the respondents were encouraged by reviewers, editors and during their research training to employ selective reporting of hypotheses and HARKing.

To summarize, about half of the survey participants admitted to selective and misleading reporting of research outcomes, and about a third reported selectively including and excluding data in order to get statistically significant results—all in studies published between 2009 and 2013 in management journals (see Banks et al. 2016, Appendix A and B). When journal articles are scrutinized, more evidence is found that research that was reported in such a questionable manner gets published. For example, Bergh et al. (2017) subjected articles in the *Academy of Management Learning and Education* to three tests and identified several red flags suggesting that findings were dishonestly reported. Similar findings have been observed in other academic disciplines as well (Head et al. 2015). Even more troubling is the observation that methodology teachers, reviewers and editors do recommend some of these practices suggesting that they have become a norm in business research which competes with the norms endorsed by more scrupulous academics (Banks et al. 2016).

Publication Bias

The question of whether a publication bias against null findings and in favour of significant p values really exists is crucial for ensuring that published research can be trusted on the terms of a quantitative community. A wide number of commentaries and studies have addressed this problem across the physical and social sciences (e.g. Ferguson and Heene 2012), typically in relation to the 'file-drawer problem' discussed in the meta-analytic literature (see Rothstein et al. 2006; see also <http://psychfiledrawer.org/>). Overwhelming evidence shows that research in which the

focal hypothesis was rejected due to a high p value has a smaller likelihood of being published (Greenwald 1975; Hopewell et al. 2009). For example, Franco et al. (2014) found that roughly 21% of a set of 221 of nationally representative studies with null findings were published, versus 62% for those with statistically significant findings. This profound difference of 40% points, however, is not the entire story, because the authors also found a 60% points difference in the rates at which researchers bother to write up papers reporting null versus statistically significant findings, providing evidence that researchers are fully aware of this publication bias. Moreover, a study of three marketing journals observed that the percentage of published studies wherein no support for the focal hypothesis was found has declined by one-half between the 1970s and the 1980s, suggesting a trend towards more publication bias over time (Hubbard and Armstrong 1992).

Supporting these findings, researchers and Ph.D. candidates have concerns that top journals are less likely to publish an article containing too many null findings (Banks et al. 2016). In the absence of a reliable and comprehensive study of management journals, we can only speculate to what extent a publication bias exists and how it manifests (although see Kepes et al. 2012). The first suspects are of course editors and reviewers who prefer studies in which support for new theoretical insights is reported over studies which question the validity of existing theory with null findings (Cortina and Folger 1998). After all, assigned action editors and reviewers (as experts in the field) are likely to have contributed to the establishment of studied theories and may therefore have a stake in the evaluation of their relevance.

On the other hand, the high percentage of researchers admitting to dishonest reporting practices and selectively writing up significant findings suggests that authors actively censor their results in anticipation of biased reviewers and editors. Additional evidence for such practices was found in a study comparing PhD dissertations and journal articles based on these dissertations. O'Boyle et al. (2017) found that the ratio of supported to unsupported hypotheses more than doubled between Ph.D. defence and publication. Practices such as the deletion or addition of data, variables and hypotheses were observed, but were not reported in published manuscripts. Also, many of these articles were published with members of the Ph.D. committee, suggesting that these more experienced researchers not only condone but may teach and endorse the practices (see also Bettis 2012). This should not be surprising, in part because researchers have a wide array of potential ways of achieving statistical significance in their research, allowing almost anything to appear as statistically significant with enough time, effort, expertise and ingenuity (see Simmons et al. 2011).

Bias Against Evidence that Questions a Theory's Validity

There are several reasons why few researchers publish evidence that fails to support a theory—beyond the fact that typical null hypothesis significance tests were not designed and are not suited to measure evidence in favour of a null hypothesis (see Berger and Sellke 1987; Zyphur and Oswald 2015). First, there is the vested interest of authors, reviewers and editors to demonstrate that their field of research is based on empirically adequate theory. In business research, we rarely see that a new theory completely replaces an older one with the kind of incommensurability described by, for example, Kuhn (2012). Often, new propositions are framed as extensions or boundary conditions of existing theories. Few business researchers take the risk Giordano Bruno took when he proposed to replace a Ptolemaic model of the universe with a heliocentric one. Although he had empirical evidence on his side, he was burned for heresy by the church (rather than colleagues or a tenure committee). As a result, empirical results that challenge theories and ideas that culturally engrained (e.g. good behaviour will be rewarded) or rooted in a long academic tradition are likely to meet more scepticism in the reviewing process. Consequently, by claiming to make incremental improvements to theory, academics avoid hostility while keeping on the good side of researchers who they are likely to encounter as reviewers, editors or collaborators. However, major innovations to business theory and a pruning of our body of theories cannot be expected when researchers decline to write up and editors do not publish null findings.

A second reason why most empirical research is presented as supporting theory is that many researchers have too much confidence in p values, frequently interpreting a p value under 0.05 as evidence that a studied relationship is reliable and therefore trustworthy. Methodologists have long warned that a significant p value should be interpreted as reason to continue studying a relationship, but not as proof of its validity in a general sense (Gigerenzer and Marewski 2015; Head et al. 2015; Johnson 2013). In 2016, the American Statistical Association published a statement on how to interpret and report p values (Wasserstein and Lazar 2016), which cautions about their use—although without adequate critical insight about questionable research and reporting practices.

Several attempts have been made to probabilistically establish the conditional probability of failing to support a hypothesis in a sample of data when the hypothesis is 'true' in a larger population (of course, within the epistemological constraints of a reality wherein hypotheses and theories may be either 'true' or 'false' in abstract, unobserved populations). This is typically investigated by probabilistically calculating how often a hypothesis may be true of a population

under the condition that it is rejected with a p value below 0.05 in a sample (Berger and Sellke 1987). Assuming three classes of distributions of the tested variable (symmetric, unimodal and symmetric, normal), relative probabilities between 12.8 and 32.1% were calculated for a hypothesis submitted to a single test (Berger and Sellke 1987). Under the assumption of weak evidence such as a small effect size, “the probability of getting a p value near 0.05, when H_1 is true, cannot be much bigger than the probability of getting a p value near 0.05, when H_0 is true” (Sellke et al. 2001, p. 64). Employing an alternative estimation method, Johnson (2013, p. 5) concluded that ‘between 17 and 25% of marginally significant scientific findings are false’. In particular, implausible hypotheses with low prior odds are more likely to result in ‘false alarms’, as for p values at 0.5, a stunning 89% of the findings are not referring to a ‘real effect’ (Nuzzo 2014, p. 151). In contrast, findings for a hypothesis with good prior odds are estimated to result in only 4% of such false alarms (Nuzzo 2014). In short, in situations in which researchers face the most difficulties to make a good judgement call regarding whether a hypothesis is trustworthy or not, the p value is losing its power to give us guidance.

This type of statistical error is likely to be one of the factors contributing to the low success rate of replication although other factors such as sampling errors contribute to this issue as well. For example, an analysis of 100 replication studies of 98 original psychological studies resulted in only 39 unambiguously successful replications (Open Science Collaboration 2015). Despite these discouraging numbers and the fact that the relative probability of a hypothesis being ‘true’ may increase if it was statistically supported in two independent studies, few replication studies are conducted and even fewer published. Instead, researchers prefer to test new hypotheses in order to make incremental improvements to theory. The result is that some academics estimate that 85% of research resources across the sciences are wasted on studies that provide unreliable results (Ioannidis 2014), contributing to a crisis of trust in science.

To conclude, there are several reasons why the research we publish may not be as trustworthy as it should be. We have too much confidence that a significant p value is a strong indicator of the existence of a relationship and is therefore trustworthy. As a result, not enough replication studies are conducted and too much focus is put on developing new hypotheses rather than testing old ones to help establish their credibility. Moreover, if researchers engage in questionable research practices, the number of hypotheses that are supported in a replication study may diminish further. Given that studies indicate that a substantial proportion of statistically significant findings may be problematic on the very statistical and probabilistic terms that are used to test hypotheses (Berger and Sellke 1987; Johnson 2013; Nuzzo 2014), increasing these percentages through

selective reporting practices and other questionable activities seems irresponsible and unethical. The net result is a need for mechanisms that promote honesty and trust, such as publishing replication studies and discouraging questionable research and reporting practices by other means, which we now treat.

Publishing Null Findings, Replications and Preregistered Research at JBE

In discussions of questionable research practices among quantitative researchers, most authors identify a set of similar remedies. These include journals being open to the publication of null findings, replication studies and preregistered research in conjunction with improved methodology training and more transparency in research reporting (Banks et al. 2016; Bettis et al. 2014; Byington and Felps 2017; Schwab and Starbuck 2017). Joining this discussion on honesty and trust, we endorse these practices at JBE and note that, henceforth, JBE will be more open to publishing null findings and replication studies. In addition, JBE takes steps to test a protocol for preregistered research, which has the potential to increase transparency and honesty in the editorial and research process.

In general, the main goal of preregistering research is that hypotheses as well as a study’s design, sampling process and analysis plan are published (typically online) before data collection begins. This practice allows reviewers and editors to distinguish between a priori hypotheses and post hoc analysis, limiting the ability of researchers to dishonestly present the latter as if they were motivated by the former. A growing number of academic disciplines encourage preregistered research, and there are several websites available for authors to register their research plan. Some websites publish preregistration in a specific discipline such as the WHO Registry Network for clinical trials (<http://www.who.int/ictrp/network/en/>), the American Economic Association’s registry for randomized controlled trials (<https://www.socialscienceregistry.org/>) and the Evidence in Governance and Politics initiative (<http://egap.org/>). The Open Science Framework is open for preregistration of studies from all disciplines (<https://osf.io/>) and provides a template outlining which information should be included in the preregistration (<https://osf.io/sgrk6/>). User friendly and open to all disciplines, but less detailed is the preregistration process at AsPredicted (<https://aspredicted.org/>), where researchers provide information by answering eight questions regarding their research project. This preregistration can be done anonymously, which allows researchers to provide a link to the preregistered research plan in the methodology section of their article without compromising their anonymity in the review process (for an example, see Banks et al. 2016).

It is important to note that many research contributions do not qualify for preregistration. Due to their nature, theory-based contributions as well as exploratory research (qualitative or quantitative) are based on a more iterative research processes for which preregistration offers no benefits. Such inductive and abductive research relies on analysing data for emergent patterns in order to avoid overlooking existing relationships, which may be relevant for developing or improving theory. In order to uncover initially unknown and unexpected patterns, researchers often combine various data analysis approaches and may decide to probe further when initial evidence of a given pattern is observed. In particular, researchers may decide to collect more data or employ another analytic tool in order to learn more about a phenomenon. As a result, the initial research plan is subject to change and several rounds of data collection and analysis may be the rule rather than the exception. Thus, preregistering a research plan would hinder rather than improving exploratory research as it is usually not foreseeable what patterns will emerge and which data and analytic methodologies will be best suited for analyses.

In turn, it is important to distinguish exploratory and hypothesis-testing research, because questionable research practices blur the line between these two approaches such that the research is done in an exploratory fashion but is reported as being a hypothesis-testing study. When exploratory research is conducted with quantitative data, a wide range of relationships between variables are statistically evaluated. However, under common logics of probabilistic inference, the more tests are run on the same data set, the higher is the likelihood that classic Type I/II errors occur. This means that some relationships appear to be more relevant than they really are because a significant p value is observed ('false' positives) and for some relationships that are usually relevant, no significant results are found ('false' negatives), due to 'noise' or 'error' in the data. However, under this same logic, researchers can never identify which of the tested relationships are 'correctly' positive, 'false' positive, 'correctly' negative or 'false' negatives, and therefore researchers can only evaluate theory based on whatever findings are extracted from their data rather than what is abstractly 'correct' or 'false' outside of what is observed. Consequently, the story goes that any theory which is developed from exploratory research should be treated with caution as more empirical research will be needed in order to determine its reliability and trustworthiness through additional data collection.

When exploratory research is conducted in a qualitative manner, the question of reliability is on the mind of the researcher and will be expressed by discussing the transferability and generalizability of the findings. Consequently, qualitative researchers often recommend further studies of the relationships they described. Developing theory from

exploratory quantitative research is in some aspects similar to that in exploratory qualitative research. The researcher has to ask the question whether the relationships which were observed in their study are likely to be transferable to other (similar) circumstances and likely to be stable over time. Researchers make judgement calls when they decide which findings from their study are relevant for theory testing (Kuhn 2012). This judgement call will be related to how strongly the data support an inference as well as how persuasive the associated theory is in the context of other theories. If it is not possible to explain how one variable is influencing another, a finding may not be deemed trustworthy and researchers may decide to report the finding without using it for subsequent theory development. However, some doubt will always remain in terms of whether findings are interpreted correctly. Thus, the practice to conducting more studies including hypothesis-testing studies on the same phenomenon can help determine which theories are trustworthy, independent from the researcher studying it and the methods used—such that science is a collective, democratic endeavour within a community with trust being paramount to its conduct (Peirce 1923). The point is that by evaluating multiple studies on what is deemed to be the same phenomenon, including replication research, it is possible to find out whether a theory describes a pattern that is stable over time and observable in different situations.

At JBE, we encourage a rich variety of research designs and methods, as we believe that they can complement each other in many cases. Therefore, submissions of conceptual, qualitative and quantitative exploratory research and hypothesis-testing research are all welcomed, including studies that present null findings and serve as replications of past research. In order to foster the idea of preregistration and learn about its merits and possible pitfalls, we are issuing a call for contributions for a Special Issue at the *Journal of Business Ethics*.

Special Issue on Preregistered Research

We invite two types of contributions to the Special Issue: studies following the preregistration protocol testing business ethics theories as well as contributions that critically evaluate hypothesis testing and the preregistration process. Contributions following the preregistration protocol can, for example, test business ethics theories and propositions for which the available empirical evidence is inconclusive and/or have not yet been subjected to systematic replications. We also welcome studies that aim to provide direct replications of prior studies, for example in the case of experimental research, as well as conceptual replications in which the context of the study, such as place or industry, varies from the original study (for example, in case of field research,

Lynch et al. 2015). Moreover, we welcome studies that aim to test hypotheses for the first time based on well-reasoned theoretical assumptions or exploratory research findings. In addition, we invite contributions that critically evaluate the epistemological and ontological foundations of hypothesis testing as well as the preregistration process. More generally, we call for papers that take a critical approach to the ethics, practices and logics that create dilemmas of trust and uncertainty which motivate preregistration, as well as the kind of hypothetico-deductivism that researchers appreciate when probabilistically quantifying uncertainty with the goal of maximizing the validity of statistical inferences.

The first type of contributions to the Special Issue follows a review protocol for preregistered research for all empirical contributions. So far, 49 journals, mostly in the field of psychology and neuroscience, have introduced preregistered reporting protocols (Centre for Open Science 2017). There are two main differences between this review protocol and the preregistration process described above. Firstly, researchers submit their research plan directly to JBE, rather than making it public in a registry. Secondly, reviewers and editors evaluate a research plan, not a manuscript describing the finished study. Research plans are evaluated based on their potential to test hypotheses relevant to business ethics theories and their contribution to examining moral aspects of systems of production, consumption, marketing, advertising, social and economic accounting, labour relations, public relations, organizational behaviour and related topics. Data collection and analysis take place after a study is initially accepted for publication.

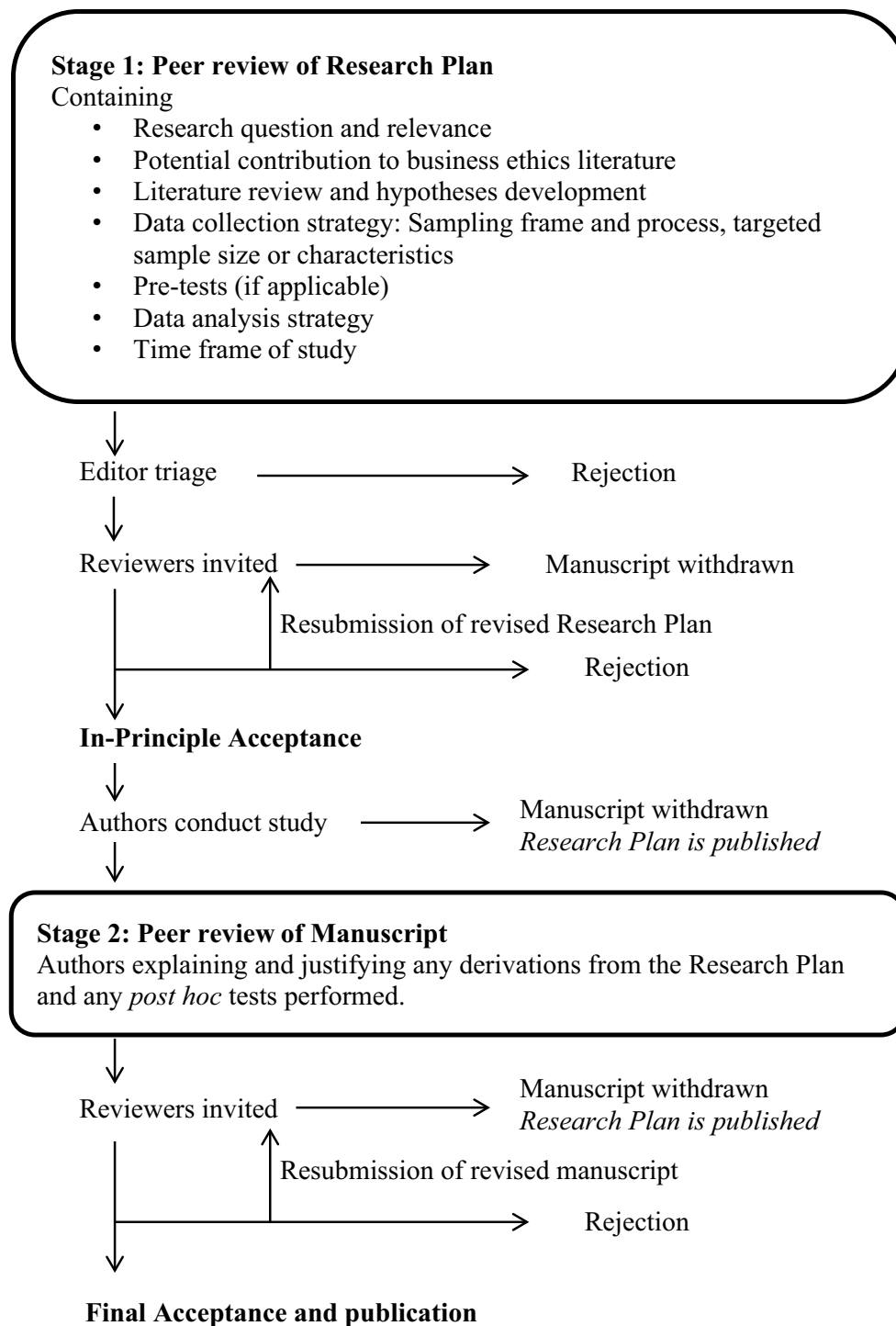
The advantage of this approach is that researchers receive reviewer feedback early in the research process when it is still possible to make changes to the study without compromising the integrity of the research process in terms of trust (as when reviewers recommend HARKing or dropping some hypothesis tests in a submitted manuscripts). Thus, recommendations regarding relevant literature, alternative scales for measuring constructs, more appropriate ways of data collection and conducting analyses are offered before data collection and analysis start. If a research plan is rejected, the researchers can decide whether they still want to collect data on the basis of the same or a revised research plan for submission elsewhere or whether to abandon the project. By getting early feedback from their peers—although increasing the reviewing and editorial burden—researchers are able to focus their resources on those projects that are most promising in terms of peer opinions.

Figure 1 outlines the preregistered research process we will follow for the Special Issue. In order to preregister a study, a research plan is submitted containing a detailed abstract, introduction, literature review, hypotheses development, a detailed plan for data collection and analysis and, if applicable, a description of pre-testing procedures and pilot

data, as well as a proposed time frame. Attention should be paid to the question whether the sample resulting from the data collection strategy is likely to be in terms of quality and size adequate to study the proposed relationships. At the time of submission, authors consent to the publication of the research plan, if the plan is accepted in principle, but the study is withdrawn afterwards. Further, all authors declare that they are aware of the requirement that data collection is only permitted after the study has been accepted.

After a formal screening by the editor, the research plan is sent out to reviewers or desk rejected if it does not fit at JBE (e.g. if it does not describe a hypotheses testing study, if it is too weak to justify reviewing or irrelevant to the business ethics discourse). As with other manuscripts, reviewers can recommend accepting, revising or rejecting a study. At this stage, all revisions are made before data collection starts and can address, for example, inclusion of literature, rephrasing of hypotheses and improvements on data collection and analysis strategy. Reviewers are encouraged to pay attention to the potential of the study to make a contribution to business ethics theory and to deliver meaningful and statistically sufficiently powerful evidence. Once the study's design is approved by reviewers, the editor accepts the research plan in principle. An in-principle acceptance indicates that a study following the agreed-upon research plan will be published irrespective of its results. Thus, even if no hypothesis is supported, the results will be published under the condition that they are presented and discussed in an appropriate manner—meeting the quality standards of JBE. Based on the time frame proposed by the authors in the research plan, the editor sets a deadline for re-submission. As this is a Special Issue, we can only accept studies for which data collection and analysis are feasible within a period of 9 months.

Once the proposal is initially accepted, the authors collect and analyse the data and, finally, submit a complete manuscript. This is once more reviewed, preferably by the same reviewers who evaluated the research plan. The main focus of reviews during the second stage is to determine whether the research plan was followed and if any changes which had been made by the authors to the research process were appropriate and do not compromise the quality of the study. For example, the sample size of the final study might be smaller than the targeted sample size. Or a scale used in the study failed to produce reliable results, making it infeasible to follow the original data analysis plan. The reviewers are asked to carefully assess on the basis of the information provided by the authors in the manuscript and an explanatory letter to the reviewers and the editor, whether the reported changes compromise the integrity of the study or not and whether the manuscripts meets JBE's standards of quality and scope. In order to improve the quality of the manuscript, reviewers can require revisions of the presentation of the data analysis, discussion and conclusions. For example,



Adapted from Chambers (2014)

Fig. 1 Submission and review process of preregistered research at the Journal of Business Ethics. [Adapted from Chambers (2014)]

additional post hoc analysis can be suggested, if the results hint at a relationship which was not covered by the original analysis plan such as unexpected interactions between variables. Reviewers can also propose the discussion of

alternative interpretations of the results and their implications for theory and practice.

Given that the study was conducted and analysed in line with the initial design and meets the quality standards of

JBE, the manuscript will be accepted for publication in the Special Issue. Rejections at this final stage are limited to manuscripts in which:

- a. The approved study design was not implemented. Any deviations from the research plan need to be explained and justified in a letter to the reviewers and editor;
- b. If the time needed for data collection exceeds the deadline set for the Special Issue;
- c. Do not meet the quality standards of JBE after a final revision.

To be clear, rejection is not possible on the basis of null findings; JBE is committed to publish all results (including, as we mentioned, in papers that are not preregistered, although obviously such studies do not benefit from pre-review and acceptance as in the preregistration protocol).

The protocol outlined above results in a number of challenges for authors, reviewers and editors as well as the academic community and its related institutions more generally. Researchers must plan ahead in order to develop a detailed and feasible research plan for a theoretically relevant and methodologically sound study. Reviewers must be able to anticipate which problems can arise in the outlined study in order to evaluate its potential to make a relevant contribution to knowledge and to propose improvements. Editors must ensure that authors and reviewers understand the process well and that sufficient time is provided for data collection, analysis and the writing of the final manuscript. The academic community is challenged to embrace the fact that null findings and replications are relevant for knowledge creation and therefore worth reading and being cited. Academic institutions must support journals and researchers who commit to improving research practices, even if old and valued indicators for academic success such as impact factors and numbers of publications may need to be re-evaluated in this process. This raises the question why we should engage in such a cumbersome experiment such as the testing of a new peer review protocol. The following section aims to answer this question by analysing the potential benefits of this editorial protocol.

How Does Preregistered Research Encourage Trust?

There is no single remedy for ensuring trust in the entirety of a research process and researchers. However, we believe that preregistered research sets incentives for conducting research that can increase honesty and trust. Firstly, this new approach encourages researchers to screen the empirical evidence provided to support business ethics theories and encourages them to identify areas in which hypothesis

testing and replication research are still needed. This practice should discourage the perpetuation of propositions unwarranted by observation (Harzing 2016).

Secondly, it encourages the development and peer review of a research plan. By providing feedback early in the research process, errors can be avoided and better research plans are developed. For example, studies based on ambiguous hypotheses, questionable scales and insufficient samples are avoided.

Thirdly, researchers have no reason to selectively report results and to hide null findings, as JBE is committed to publish all results of the proposed analysis. Therefore, questionable research practices such as selective inclusion of data or selective reporting of findings do not increase the chance of publication. Finally, the presentation of post hoc findings as a priori hypotheses is not feasible in this process.

Overall, the most commonly used questionable research and reporting practices do not offer any benefit to authors in the preregistered research process. Successful studies are most likely to be theoretically relevant, carefully planned, well executed, reported in a transparent manner and critically evaluating all findings and their implications for theory development. We expect publications on preregistered research will provide insights regarding which theoretical assumptions are strongly supported in various empirical situations and for which relationships we are only occasionally able to find empirical support.

Ultimately, a research tradition in which exploratory research, theory development and hypotheses testing are informing each other should produce: a set of highly reliable theories; others that need refinement as their validity may be limited to a specific context; and some theoretical assumptions that have to be dismissed as empirical findings show no systematic support for them. Statistical meta-analysis also become more meaningful when null findings are systematically reported—facilitating, at least, public access to these studies for conducting meta-analyses. Consequently, our ability to predict on the basis of our theories which managerial interventions are most likely to be beneficial and successful may increase—conditional on, of course, highly contextual reasoning regarding which approaches to intervention may be best in a given problematic situation. In sum, all of this is meant to increase the trust in researchers and their results that is so vital for the creation and maintenance of research communities.

Compliance with Ethical Standards

Conflict of interest Both authors declare no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Banks, G., O'Boyle, E. H., Jr., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., et al. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, *42*(1), 5–20.
- Berger, J. O., & Sellke, T. (1987). Testing a null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, *82*(397), 112–122.
- Bergh, D., Sharp, B., & Li, M. (2017). Tests for identifying “red flags” in empirical findings: Demonstration and recommendations for authors, reviewers and editors. *Academy of Management Learning and Education*, *16*(1), 110–124.
- Bettis, R. A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal*, *33*, 108–113.
- Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. (2014). Quantitative empirical analysis in strategic management. *Strategic Management Journal*, *35*, 949–953.
- Byington, E., & Felps, W. (2017). Solutions to credibility crisis in management science. *Academy of Management Learning and Education*, *16*(1), 142–162.
- Centre for Open Science. (2017). *Registered Reports: Peer review before results are known to align scientific values and practices*. Available online under: https://cos.io/rr/?_ga=1.103210176.1532854806.1489421591. Accessed 4 April 2018.
- Chambers, C. (2014). *Registered reports: A step change in scientific publishing*. Available online under: <https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/registered-reports-a-step-change-in-scientific-publishing>. Accessed 4 April 2018.
- Community for Responsible Research in Business and Management. (2017). *A vision of responsible research in business and management: Striving for useful and credible knowledge*. Position Paper published online under: http://rrbm.network/wp-content/uploads/2017/11/Position_Paper.pdf. Accessed 4 April 2018.
- Cortina, J. M., & Folger, R. G. (1998). When is it acceptable to accept a null hypothesis: No way, Jose? *Organizational Research Methods*, *1*, 334–350.
- Cortina, J. M., & Landis, R. S. (2011). The earth is not round ($p = .00$). *Organizational Research Methods*, *14*, 332–349.
- Cumming, G. (2014). The new statistics why and how. *Psychological Science*, *25*, 7–29.
- Dewey, J. (1920). *Reconstruction in philosophy*. New York: Holt Publishing.
- Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Holt Publishing.
- Du Gay, P. (2015). Organization (theory) as a way of life. *Journal of Cultural Economy*, *8*(4), 399–417.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904.
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555–561.
- Fish, S. (1985). Consequences. *Critical Inquiry*, *11*, 433–458.
- Fish, S. (2003). Truth but no consequences: Why philosophy doesn't matter. *Critical Inquiry*, *29*, 389–417.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*, 1502–1505.
- Gigerenzer, G., & Marewski, J. N. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, *41*(2), 421–440.
- Greenwald, A. G. (1975). Consequences of prejudice against the Null hypothesis. *Psychological Bulletin*, *82*(1), 1–20.
- Harzing, A.-W. (2016). Why replication studies are essential: Learning from failure and success. *Cross Cultural and Strategic Management*, *23*(4), 563–568.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of P-hacking in science. *PLoS Biology*, *13*(3), e1002106.
- Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*, *1*, MR000006.
- Hubbard, R., & Armstrong, J. S. (1992). Are null results becoming an endangered species in marketing? *Marketing Letters*, *3*(2), 127–136.
- Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine*, *14*(10), e1001747.
- Jasanoff, S. (Ed.). (2004). *States of knowledge: The co-production of science and the social order*. New York: Routledge.
- Jasanoff, S. (2009). *The fifth branch: Science advisers as policymakers*. Cambridge: Harvard University Press.
- Jasanoff, S. (2010). Testing time for climate science. *Science*, *328*, 695–696.
- Jasanoff, S. (2014). A mirror for science. *Public Understanding of Science*, *23*, 21–26.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.
- Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, *110*(48), 19313–19317.
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, *15*, 624–662.
- Knorr-Cetina, K. D. (2009). *Epistemic cultures: How the sciences make knowledge*. Cambridge: Harvard University Press.
- Knorr-Cetina, K. D. (2013). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. New York: Elsevier.
- Kuhn, T. S. (2012). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. *Science, Technology and Human Values*, *42*(2), 191–202.
- Lynch, J. G., Jr., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing*, *32*, 333–342.
- Morey, R. D., Rouder, J. N., Verhagen, J., & Wagenmakers, E. J. (2014). Why hypothesis tests are essential for psychological science a comment on Cumming. *Psychological Science*, *25*, 1289–1290.
- Nuzzo, R. (2014). Statistical errors. *Nature*, *506*(7487), 150–152.
- O'Boyle, E. H., Jr., Banks, G. C., & Gonzalez-Mule, E. (2017). The chrysalis effect: How ugly initial results metamorphose into beautiful articles. *Journal of Management*, *43*(2), 376–399.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943.
- Peirce, C. S. (1923). *Chance, love, and logic: Philosophical essays*. London: Kegan Paul, Trench, Tubner and Co. LTD.
- Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. Chicago: University of Chicago Press.
- Porter, T. M. (1986). *The rise of statistical thinking, 1820–1900*. Princeton University Press.
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. New Haven: Princeton University Press.

- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. New York: Wiley.
- Schwab, A., & Starbuck, W. H. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning and Education*, 16(1), 125–141.
- Sellke, T., Bayarri, M. T., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71.
- Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England*. Chicago: University of Chicago Press.
- Shapin, S. (2009). *The scientific life: A moral history of a late modern vocation*. Chicago: University of Chicago Press.
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton: Princeton University Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Starbuck, W. H. (2016). 60th anniversary essay: How journals could improve research practices in social sciences. *Administrative Science Quarterly*, 61(2), 165–183.
- Strathern, M. (2000). The tyranny of transparency. *British Educational Research Journal*, 26(3), 309–321.
- Tsoukas, H. (1997). The tyranny of light: The temptations and the paradoxes of the information society. *Futures*, 29(9), 827–843.
- Van Fraassen, B. C. (2008). *The empirical stance*. New Haven: Yale University Press.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129–133.
- Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management*, 41, 390–420.