On Not Confusing the Tree of Trustworthy Statistics with the Greater Forest of Good Science: A Comment on Simmons et al.'s Perspective on Preregistration

Article *in* Journal of Consumer Psychology · February 2021

DOI: 10.1002/jcpy.1213

2 authors:

Michel Tuan Pham
Columbia University

Travis Tae Oh
Yeshiva University

On Not Confusing the Tree of Trustworthy Statistics with the Greater Forest of Good Science: A Comment on Simmons et al.'s Perspective on Preregistration

Michel Tuan Pham[1] and Travis Tae Oh[2]

[1]Columbia University

[2]Yeshiva University

Too appear as a commentary as part of a Research Dialogue on Preregistration in the *Journal of Consumer Psychology*, Vol. 21 (1), January 2021 (https://doi.org/10.1002/jcpy.1213). A related commentary article, titled "Preregistration Is Neither Sufficient nor Necessary for Good Science," appears in the same issue (https://doi.org/10.1002/jcpy.1209).

Abstract

In this commentary on Simmons, Nelson, and Simonsohn (this issue), we examine their rationale for preregistration within the broader perspective of what good science is. We agree that there is potential benefit in a system of preregistration if implemented selectively. However, we believe that other tools of open science such as the full sharing of study materials and open access to underlying data, provide most of the same benefits—and more (i.e., the prevention of outright fraud)—without risking the potential adverse consequences of a system of preregistration. This is why we favor these other means of controlling type-I error and fostering transparency. Direct replication, as opposed to conceptual replication, should be encouraged as well.

In their target article, Simmons, Nelson, and Simonsohn (SNS, this issue) summarize the theoretical rationale for the use of preregistration as a scientific practice and provide some practical recommendations on how to make preregistrations most effective. By and large, the arguments that SNS summarize in their article parallel those advanced in previous articles championing the practice of preregistration (e.g., Nosek et al., 2018; van't Veer and Giner-Sorolla, 2016). However, whereas prior discussions of preregistration focused on its contribution to transparency and open science, SNS place a relatively greater emphasis on the benefits of preregistration with respect to preventing "p-hacking" in order to increase confidence in empirical findings.

As stated in our own target article (Pham & Oh, this issue), we fundamentally support the Open Science movement's goal of fostering greater transparency and reproducibility of scientific findings. Hence, we fully agree with SNS's view that social sciences—like all other sciences—need to be based on correct facts. To this end, we concur that effective means of curbing the practice of p-hacking are needed, and, as conveyed in our target article, we do see some merit in certain uses of preregistration (e.g., verifying the effect size and boundary conditions of research findings for business and policy applications; setting a higher bar for research that challenges well-established scientific beliefs).

Notwithstanding our reservations about preregistration as a scientific norm, we additionally find SNS's practical guidelines on the implementation of preregistration quite useful. As these authors explain, on the AsPredicted.org website, preregistrations involve submitting answers to a small number of specific questions. In their target article, SNS provide examples of good and bad answers to these questions, along with a brief explanation (see their Table 1). We find these examples to be very helpful, and we agree that they ease the procedural cost of preregistering. The checklist that SNS provide to both researchers *and reviewers* is also helpful. As we noted in our target article, a proper system of preregistration requires a complementary system of monitoring, which SNS suggest should be handled by journal reviewers; they offer specific recommendations on how reviewers should evaluate the conformance of journal submissions with the corresponding preregistrations. These suggestions make sense.

Where our and SNS's positions on preregistration differ is that SNS focus primarily on how preregistration can help reduce p-hacking, thereby increasing the trustworthiness of test statistics, whereas we consider preregistration from a broader perspective of fostering better science. From this larger perspective, limitations, opportunity costs, and adverse effects of preregistration that are not otherwise obvious become more apparent. It is these issues that make us more hesitant about an unconditional embrace of preregistration as a scientific norm.

What is Good Science?

As illustrated in Figure 1, we believe that good science is a function of (a) the tools that the scientists use; (b) the epistemological criteria that the research meets; (c) core qualities of the scientific outputs produced; and (d) how well those scientific outputs help fulfill broader societal goals.

The Tools of Science

Good science is not just the appropriate use of statistics or the design of clean experiments. It involves a whole range of processes and tools that contribute to the eventual production of trustworthy and useful scientific evidence. One of these is keen observation, which facilitates the detection of interesting phenomena that are worthy of further investigation.

Another tool is exploration, which the inquisitive scientist uses to probe things to unearth what lies below the surface. A third tool is a formal hypothesis, which crystallizes the scientist's core prediction and intended contribution. A fourth tool is a strong study design and set of procedures, which enable a diagnostic test of the researcher's hypothesis.

Another essential component of good science is a rigorous analysis of the data: Is the analysis appropriate for the data at hand and performed correctly? Are the conclusions drawn from the analysis accurate? Following the analysis, a proper reporting of the research—its objectives, methodology, results, and conclusions—would be accurate, clear, and complete.

The quality of science also depends on the direct replication of the studies by the researchers themselves, which in our opinion is not performed often enough. Such a practice, if more widely adopted, would go a long way toward reducing the rate of type-I error, the primary motivation behind SNS's embrace of preregistration. Conceptual replication is also helpful, especially for evaluating the generalizability of empirical findings. However, contrary to what some authors have suggested (Lynch et al., 2015), it is not a substitute for direct replication (see Pashler & Harris, 2012).

The next set of tools is more institutional. Obviously, an effective peer-review process plays an important role in the quality of science. Ideally, the review process should balance the risk of type-I and type-II error. However, major journals tend to put relatively greater emphasis on the control of type-I error, which is also SNS's primary focus. We believe that, somewhat paradoxically, it is the very focus on the control of type-I error with a nominal threshold of $p < .05$ that has contributed to the problems of p-hacking that SNS have pointed out in their influential work (e.g., Simmons, Nelson, & Simonsohn, 2011; Simonsohn, Nelson, and Simmons, 2014).

The Open Science movement has championed another set of tools that are meant to promote the transparency and reproducibility of science (Nosek et al., 2015). These tools are identified with dashed boxes in Figure 1. They include (a) the full sharing of methodological and analytical details of each study (stimuli, instruments, programming codes, etc.); (b) open access to the study data; (c) the independent replication of published research; and (d) the preregistration of studies, which is the focus of this dialogue and SNS's article. As shown in Figure 1, these Open Science tools and the more classic tools described earlier help support a variety of epistemological ideals that define good science.

Scientific Ideals

Good science is defined not just by the tools that it employs but also by the epistemological ideals that guide the research. At the very heart of these ideals is a fundamental principle of veridicality. Science must be about verifiable truths. When SNS describe the job of the scientist as involving (a) the discovery of true facts about the world, and (b) interpreting those facts for theories, they are referring to this principle. Veridicality entails the controlling of type-I error, so that false results are not interpreted as "true," which is SNS's primary concern. However, we would argue that the principle of veridicality also entails the control of type-II error, so that true results are not disregarded as "false." In addition to balancing type-I and type-II errors, the principle of veridicality assumes the integrity of data, which presumes the absence of fraud. A final aspect of veridicality is the robustness of data patterns across minor variations in methodology that are not theoretically meaningful (e.g., the use of slightly different stimuli, alternative definitions of outliers, different batches of respondents across experiments).

Other ideals of good science involve transparency, which is a primary goal of the Open Science movement, and generalizability, without which science would lose much of its purpose.

Empirical findings that are not generalizable beyond the limited confines of the setting where they were observed (e.g., consulting projects) are of limited scientific value.

Although the following are not always considered when evaluating science, we believe that good science additionally depends on two principles: (a) novelty/insight and (b) relevance. Good science does not just uncover facts that are true; it also uncovers facts that teach us something that (a) we did not already know, and (b) is relevant and meaningful to some external constituents. In fact, as we suggested in our target article, the main problem within our field may not be the undetected presence of false-positive results but the considerable proportion of research with limited relevance (Pham, 2013; see also Inman, 2012).

Scientific Outputs

The quality of science further depends on the quality of outputs that it produces. The main goal of science is to provide evidence that is trustworthy, or (as SNS refer to this) "true facts." On top of this, we believe that *good* science produces evidence that, in addition to being trustworthy, is *useful*. The discovery of a true fact that is not useful is not a good use of science. In other words, not all evidence that is trustworthy is necessary useful, although only evidence that is trustworthy can potentially be useful.

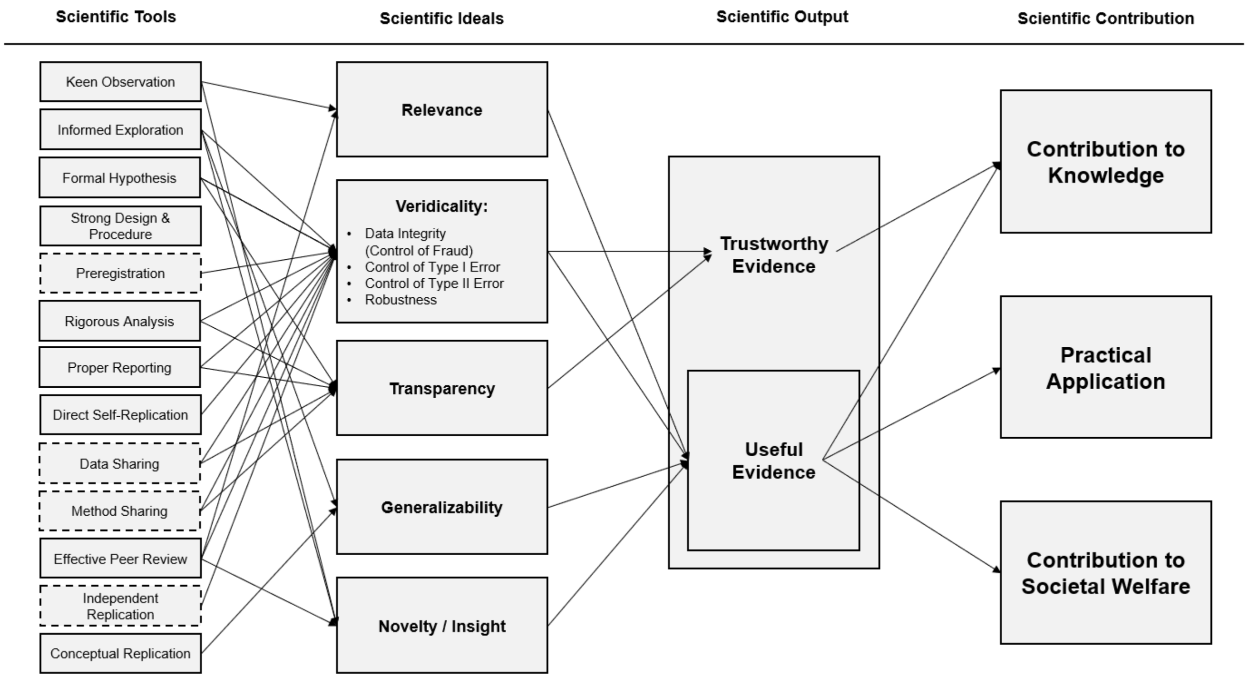

Figure 1. The Big Picture of Good Science. Note: Solid lines denote a positive impact. Dashed boxes denote major Open Science tools.

Scientific Contribution

Eventually, science should be judged by the degree to which its outputs help fulfill broader societal goals, such as (a) enriching our knowledge and understanding of the world, (b) enabling valuable practical applications, and (c) enhancing society's welfare. Such considerations make it clear that good science does not depend solely on the trustworthiness of the evidence that it yields; it additionally depends on the usefulness of this evidence to various constituents within the broader society.

Assessing Preregistration within the Bigger Picture of Good Science

Now that we have mapped out the full picture of what good science is, let us review the role that preregistration plays within this big picture. In theory, as explained by SNS, preregistration's primary function in science is two-fold: (a) to control for type-I error by preventing researchers from engaging in p-hacking practices, and (b) to increase transparency by maintaining a publicly accessible repository of preregistrations (see Figure 2). However, as SNS point out, this does not prevent outright fraud through data fabrication, nor does it prevent the deceptive practice of preregistering after results are known. Interestingly, other open-science practices such as the full sharing of study materials and open access to the study data may be more effective in terms of fraud prevention, while also providing significant (though not perfect) protection against p-hacking and false-positive results.

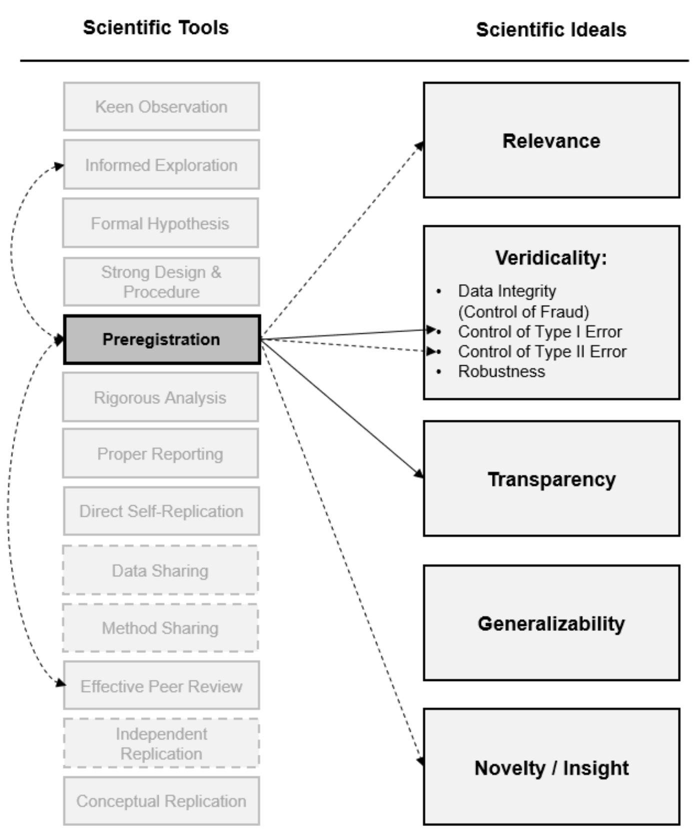

Figure 2. Preregistration's influence on other scientific tools and ideals. Note: A solid line denotes a positive impact, whereas a dotted line denotes the possibility of a negative impact.

As SNS point out, the preregistration of a study and its analysis does not guarantee that a study is not confounded or that the specified analysis is valid (which we also noted in our target article). Nor does preregistration guarantee generalizability (as pointed out in our article as well). These observations highlight the importance of not elevating preregistration above other classic scientific tools such as strong design and procedure, rigorous analysis, direct replication, conceptual replication, effective reviewing, etc. This is what we mean when we claim that preregistration is not sufficient for good science. These observations additionally illustrate the importance of considering the effects of preregistration on *other* tools and ideals of the overall scientific enterprise.

A full picture of what good science is raises a whole class of potential issues with a micro-system of preregistration, if it is implemented without a full consideration of its *total* impact on the macro-system of science (see Figure 1). Within a macro-system, any change in policy can have unintended adverse consequences. As we explain in our target article, our reservations concerning a policy of preregistration involve the risk of potential adverse effects that may eventually harm the quality of science within our field. These include (a) a likely reduction in researchers' willingness to explore (despite SNS's insistence that preregistration does not preclude exploration); (b) a significant lack of flexibility in the face of unforeseeable circumstances (see our discussion of the Field et al. [2020] study in our target article); (c) a bias toward studies that are "easy" to preregister (e.g., simple MTurk vignette studies) rather than studies that are more informative; and (d) a preference for research hypotheses that are obvious and thus more likely to be empirically supported, rather than hypotheses that are important and relevant. Such unintended consequences would be detrimental to the scientific ideals of relevance, novelty and insight, and proper control of type-II error. In addition, the fact that the preregistration of studies can serve as a heuristic signal of trustworthiness—which SNS portray as a major benefit of preregistration—increases (e) the risk of distortion of the review process if the preregistration signal is actually fake, which, as discussed in our target article, is a real possibility.

SNS argue that the mandatory registration of clinical trials in the medical and pharmaceutical field is a norm that consumer research should emulate. This analogy is misguided. Clinical medical trials are performed at the end of a very long research and development process that involves many earlier stages of preclinical research that is largely exploratory (e.g., analyses of alternate compounds, assay development, toxicology analyses). Most consumer research is much more akin to the preclinical stages of medical and pharmaceutical research than to the clinical stages that SNS encourage the field to emulate. Research that is genuinely intended for managerial application or policy intervention would be more comparable to the type of research that warrant clinical trials. And for those, we do support preregistration, as stated in our target article.

To conclude, we agree that there is potential benefit in a system of preregistration if implemented selectively (e.g., testing the effect size and boundary conditions of research findings for business and policy applications). However, we believe that other open-science tools, such as the full sharing of study materials and open access to underlying data, provide most of the same benefits—and more (i.e., the prevention of outright fraud)—without risking the potential adverse consequences of a system of preregistration. This is why we favor these other means of controlling type-I error and fostering transparency. Direct replication, as opposed to conceptual replication, should be encouraged as well.

REFERENCES

Field, S. M., Wagenmakers, E.-J., Kiers, H. A. L., Hoekstra, R., Ernst, A. F., and Ravenzwaaij, D. V. (2020). The effect of preregistration on trust in empirical research findings: Results of a registered report. *Royal Society Open Science, 7,* 181351.

Inman, J. J. (2012). Presidential address: The elephant not in the room—The need for useful, actionable insights in behavioral research. *Association for Consumer Research* conference, Vancouver, Canada, October.

Lynch, J. G., Bradlow, E. T., Huber, J. C., and Lehmann D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing, 32*(4), 333–342.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., …Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606.

Pashler, H., and Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531–536.

Pham, M. T. (2013). The seven sins of consumer psychology. *Journal of Consumer Psychology, 23*(4), 411–423.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (this issue). Pre-registration: Why and How. *Journal of Consumer Psychology*, Vol. 31 (1).

van't Veer, A. E., and Giner-Sorolla, R. (2016). Pre-registration in social psychology—a discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.