Publish less, read more: Replies to Clegg, Wiggins, and Ostenson; and to Trafimow

R. Hans Phaf
University of Amsterdam

Abstract
Clegg, Wiggins, and Ostenson, and Trafimow wrote two very different comments on “Publish Less, Read More.” In my reply to Clegg and colleagues, I agree that the inability to predict future success has led funding agencies and hiring committees to rely on social and political selection criteria and to use calculative audits as a proxy for scientific content. I argue that if it is clear that decisions to publish are based on theoretical criteria, and the numbers of publications per researcher decline, their relative value will increase, as will that of funding agencies and hiring committees. Trafimow argues that there are more than sufficient data for ad-hoc theorizing but that it does not happen, at least not enough. I agree that experimental psychologists often perform and publish research while being ill-prepared, and argue that only publication, but not data collection, should be limited to theoretically informed research.

Keywords
evolutionary dynamics, hyper-publication, social constraints, theoretical competition, unpredictability

Social constraints compete with science: Reply to Clegg, Wiggins, and Ostenson (2020)
I wouldn’t be productive enough for today’s academic system.
—Peter Higgs, 2013 Nobel Laureate in Physics (as cited in Aitkenhead, 2013)

Predicting the future, particularly in the longer term, seems to be an intrinsic impossibility. This applies not only to the weather and the stock market but also to scientific funding and hiring decisions. The inability to predict future success has led funding agencies and

Corresponding author:
R. Hans Phaf, Department of Psychology, Brain and Cognition Group, University of Amsterdam, PO Box 15915, Amsterdam, 1001 NK, The Netherlands.
Email: r.h.phaf@gmail.com
hiring committees to rely on social and political selection criteria and to use calculative audits as a proxy for scientific content. I fully agree with Clegg et al. (2020) that this represents an undesirable situation. Recently, Smaldino and McElreath (2016) showed through evolutionary modeling that these quantitative metrics result in the natural selection of poor science at the cost of the more valuable sort. Inevitably, political and social selection pressures compete with purely scientific factors, but during the last few decades the former seem to have won over the latter. Marketing skills and social “networking” increasingly matter more than the substance of one’s scientific ideas. A poignant example is the fraud case of social psychologist Diederik Stapel (Levelt et al., 2012), who received 2.2 million euro in additional subsidies from the Dutch Organization for Scientific research (NWO). At the moment, a new scientific integrity case involving hyper-publication, which also was heavily funded by NWO, is unfolding in the Netherlands (Breach of academic integrity, 2019). The funding agencies have no other option than to resort to nonscientific criteria because it is impossible to assess outstanding quality in advance and to predict future scientific success. NWO seems to be resigned to this state of affairs because it does not even evaluate the success of its subsidies afterwards.

Clegg et al. (2020) specifically address the social constraints scientists are often faced with. Although I largely agree, my concerns are about research and publication practices, particularly in experimental psychology. Still, my suggestions for improving the publication record may also help a little in restoring the balance between social and scientific factors. Let me be clear, I would not want some kind of dictatorial scheme, because I believe science thrives through its pluralistic nature. The machine-learning tool is also not meant for prescribing mandatory references but for recommending semantically related publications that may have remained hidden to authors and reviewers. I want to raise awareness that the present overpopulation of meaningless, exclusively data-oriented papers in experimental psychology has become highly problematic indeed. Theoretical coherence has been lost, as may be demonstrated by the four Problematic Publication Practices (PPPs). Eventually, the ever-increasing number of publications may also serve as a “reductio ad absurdum” of these much-eschewed production metrics (cf. Aitkenhead, 2013). When more scientific integrity cases come to light, high scores on these metrics and hyper-publication (and also researcher hyper-marketing) may come to be regarded as a sign of dubious scientific quality.

To answer the question of Clegg et al. (2020), publication is readily envisaged as the reproduction process in an evolutionary philosophy of science (EPoS), which, in different forms, was already posited by Popper (1979) and Kuhn (see Marcum, 2017). Evolutionary approaches move away from the goal of truth in a teleological philosophy of science and allow for much higher levels of variability and unpredictability. According to Marcum, Kuhn distinguished between traditional philosophy of science in which nature sets the goal for science to provide a true picture of the world and an EPoS in which science is not progressing or advancing closer to the truth but away from what is an inadequate worldview. (2017, p. 2)

Publication should aim to report present theoretical advances by showing that the research selectively strengthens one or a few of the competing hypotheses (i.e., at the cost of
“inadequate worldviews”), and should not be based on secondary data properties (e.g., preregistration, significance, etc.), or social connections. If it is clear that decisions to publish are based on theoretical criteria, and the numbers of publications per researcher decline, their relative value will increase, also for funding agencies and hiring committees.

I strongly sympathize with the idealism of Clegg et al. (2020), and would also like to see a larger commitment to human scale, not only in science but in society as a whole. I am not sure, however, that this is feasible when science expands exponentially. Even when the scientific community was much smaller, truly exceptional scientists, such as Darwin and Einstein, would not have qualified for funding prior to their great discoveries and their public recognition. I worry with Noble (2010; cf. Smaldino & McElreath, 2016) furthermore, that current assessment practices may be deeply damaging to such exceptional and mostly unconventional approaches. As Clegg et al. (2020) notice, “Developing theory is a fundamentally hermeneutic endeavor, but calculative audits (like impact factors) are precisely the opposite” (PROOF p. 3). Great discoveries are like accidents: they arise through the coincidental conjunction of a large number of serviceable factors, concerning both the social context and personal characteristics of the scientist. If only one or a few of these relatively rare factors are present, competition is relentless and will be lost, even by outstanding scientists. By the nature of this quasi-evolutionary selection process, truly innovative success can only be identified after it has happened, but not beforehand. If funding agencies want to maintain that they are able to judge the scientific future of the proposals submitted to them, they should at least start an investigation into the factors that historically were instrumental in successful research. Otherwise, they can better base their decisions on a lottery between all proposals that meet minimum standards (cf. Adam, 2019). Hume’s (1748/2008) modesty should not be limited to the just reasoner but should extend to funding agencies and hiring committees.

**Theory beats data: Reply to Trafimow (2020)**

In the fields of observation, chance only favours the mind which is prepared.

—Louis Pasteur (as cited in Vallery-Radot, 1919, p. 79)

Experimental psychologists should not only publish less but also should read more articles outside their scientific “filter bubble” (cf. Pariser, 2011). A data-oriented focus (e.g., contenting oneself with a significant “effect”) further aggravates the isolation of experimental papers from other relevant publications. Trafimow (2020) is entirely correct that there are more than sufficient data for ad-hoc theorizing but that it does not happen, at least not enough. Due to the publication overload and the statistical illusion (see Phaf, 2020), experimental psychologists often perform and publish research while being ill-prepared (cf. the above words of Pasteur, Vallery-Radot, 1919, p. 79). In my proposal, only publication, but not data collection, would be limited to theoretically informed research. Publication serves as the reproduction process in the evolutionary dynamics of science. If exclusively data-oriented research is allowed to reproduce without theory-based selection, the unceasing publication of countless unconnected papers may result in a complete scientific standstill.
Data analysis, instead of wanting to establish that something is not there (i.e., rejecting the insubstantive null hypothesis), should aim to estimate what is there and within which credible limits “this” lies. Ideally, one would compare these to theoretical estimates by formulating mathematical models that provide quantitative predictions and accuracy limits, but this does not seem very well possible in experimental psychology. Certainly, any attempt at quantifying these estimates from the data may suffer from the type of practical problems noted by Trafimow (2020). Both qualitative, narrative, reviews and quantitative meta-analyses, however, contribute to the type of meta-analytic thinking Cumming (2014) has argued for. This approach is meant to replace the notion ensuing from the statistical illusion that a result can stand on its own. Meta-analytic thinking requires extensive literature searches and looking for connections between available publications. In addition, the opportunity to gain insight into systematic differences between studies on the same topic may foster further theoretical analyses. Meta-analyses can serve as a theoretical instrument when they devise clever moderator variables that help reduce heterogeneity. Lower heterogeneity would of course point to smaller systematic differences between studies, but still cannot signal their absence. Successful moderator distinctions may thus correspond to useful hypothesis generation. As ever, even these quantifications and their associated conclusions should be taken with a grain of salt, and should only be assumed preliminary.

Although I very much appreciate Trafimow’s (2020) kind words, I am not sure whether the distinction between central and auxiliary hypotheses is all that important in an evolutionary conception of scientific development. I would argue that the distinction can only be made a posteriori when you know what the winning, and losing, hypotheses are. In my answer to his review, I made the comparison to competitive neural network models, on which I have worked a lot. A representation for a noisy stimulus (e.g., the letter “F”) consists of many different activated nodes that partially overlap with representations of other stimuli (e.g., the letters “E” and “L”). In these models, the nodes go into an excitation–inhibition battle until the lowest level of inhibition and the highest level of excitation is reached. Whether a node (e.g., representing the letter feature “_”) is auxiliary (i.e., non essential) or not then depends on which representation (e.g., the letter “E” or “F”) wins. Similarly, in the noisy evolution of science, you do not know beforehand which hypotheses will be auxiliary or not. It is even possible that auxiliary hypotheses may turn into central hypotheses later on. The suppression of the (purportedly non central) recency effect in the non replication of saccade-induced retrieval enhancement (SIRE) by Matzke et al. (2015) may present an example (see also Phaf, 2020). The activation of the to-be-retrieved material during the saccades (e.g., due to recency) would be central in my account for SIRE (see Phaf, 2017).

I much admired Trafimow’s attempts, as editor of Basic and Applied Social Psychology (Trafimow & Marks, 2015), to ban significance testing, which mirrored similar attempts of Geoffrey Loftus (1993) at Memory & Cognition in the final decade of the 20th century. I am concerned, however, that this proposal does little in combating the PPPs and in stemming the flood of over publication. The guaranteed publication of preregistered replication studies may even increase the number of meaningless papers if theoretical elaboration is not explicitly demanded by editors and reviewers.
Indeed, we do not know what counts as a successful replication. Only afterwards, when the ensemble of studies on this topic has eliminated alternative explanations in the quasi-evolutionary competition between hypotheses that meaningful research should impose, do we know whether the replication was successful or not. The almost exclusive focus on data rather than on theory, in my view, is an important cause of the so-called “reproducibility” crisis.

As I have argued in my reply to Clegg et al. (2020), the evolutionary view on scientific development entails a large degree of unpredictability. Indeed, variability is essential for any evolution to take place. For this reason, I would endorse an “anything goes” standpoint (e.g., Feyerabend, 1975), because it allows for more-or-less random mutations in the development of science. I would want to strengthen, however, the complementary selection-by-competition of hypotheses, without which random mutations will not result in meaningful progress, as may be demonstrated by the four PPPs (see Phaf, 2020). A dominant role for competition in theory building surely implies that not all hypotheses do have equal status. Some hypotheses are “fitter” (e.g., more unifying) than others, but still they cannot be guaranteed to be the fittest. Certainly, the formulation of straw-person hypotheses for enacting pseudocompetition, which Trafimow rightly condemns, may temporarily halt scientific evolution, but here, critical reviewers may help in keeping these a-theoretical papers from being published. Still, the difficulties involved must be recognized. The ideal of a linear development of science will never be fulfilled, but we will take fewer side roads, or even may avoid getting lost in the forest of research data if theoretical competition at publication is intensified.

Here I would have undoubtedly died, if not the strength of my own arm, grabbing my own pigtail, had pulled me, including my horse—which I squeezed tightly between my legs—out of it.

—(Baron Münchhausen, 2013)

**Funding**

The author received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

R. Hans Phaf [ID](https://orcid.org/0000-0002-7691-981X)

**References**

Adam, D. (2019, November 20). Science funders gamble on grant lotteries. *Nature*, 575, 574–575. https://doi.org/10.1038/d41586-019-03572-7

Aitkenhead, D. (2013, December 6). Peter Higgs: I wouldn’t be productive enough for today’s academic system. *The Guardian*. https://www.theguardian.com/science/2013/dec/06/peter-higgs-boson-academic-system

Baron Münchhausen pulls himself out of a mire by his own hair. (2013, April, 25). *Art & Popular Culture*. http://artandpopularculture.com/Escaping_from_a_swamp_by_pulling_himself_up_by_his_own_hair
The new statistics: Why and how. Psychological Science 25(1), 7–29. https://doi.org/10.1177/0956797613504966

Feyerabend, P. K. (1975). Against method: Outline of an anarchist theory of knowledge. Humanities Press.

Hume, D. (2008). An enquiry concerning human understanding. Oxford University Press. (Original work published 1748)

Levelt, W. J., Drenth, P. J. D., & Noort, E. (2012, November 28). Falende wetenschap: De frauduleuze onderzoekspraktijken van sociaal-psycholoog Diederik Stapel [Flawed science: The fraudulent research practices of social psychologist Diederik Stapel]. https://onderwijsb- rabant.nl/sites/default/files/eindrapport_stapel_nov_2012.pdf

Loftus, G. R. (1993). Editorial comment. Memory & Cognition, 21(1), 1–3. https://doi.org/10.3758/BF03211158

Marcum, J. A. (2017). Evolutionary philosophy of science: A new image of science and stance towards general philosophy of science. Philosophies, 2(4), Article 25. https://doi.org/10.3390/philosophies2040025

Matzke, D., Nieuwenhuis, S., van Rijn, H., Slugter, H. A., van der Molen, M. W., & Wagenmakers, E. J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. Journal of Experimental Psychology: General. 144(1), e1–e15. https://doi.org/10.1037/xge0000038

Noble, D. (2010). Funding the pink diamonds: A historical perspective. Notes and Records of the Royal Society 64(1), 97–102. https://doi.org/10.1098/rsnr.2009.0064

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.

Phaf, R. H. (2017). Eye movements enhance recollection of re-imagined negative words: A link between EMDR and SIRE? Journal of Experimental Psychopathology, 8(4), 364–375. https://doi.org/10.5127/jep.059916

Phaf, R. H. (2020). Publish less, read more. Theory & Psychology, 30(2), 263–285. https://doi.org/10.1177/0959354319898250

Popper, K. R. (1979). Objective knowledge: An evolutionary approach. Oxford University Press.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.

Phaf, R. H. (2017). Eye movements enhance recollection of re-imagined negative words: A link between EMDR and SIRE? Journal of Experimental Psychopathology, 8(4), 364–375. https://doi.org/10.5127/jep.059916

Phaf, R. H. (2020). Publish less, read more. Theory & Psychology, 30(2), 263–285. https://doi.org/10.1177/0959354319898250

Popper, K. R. (1979). Objective knowledge: An evolutionary approach. Oxford University Press.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society Open Science, 3(9), Article 160384. https://doi.org/10.1098/rsos.160384

Trafimow, D. (2020). Mandatory theorizing and synthesis: A comment on Phaf (2020). Theory & Psychology, 30(2), 286–291. https://doi.org/10.1177/0959354319897969

Trafimow, D., & Marks, M. (2015). Editorial. Basic and Applied Social Psychology, 37(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991

Vallery-Radot, R. (1919). The life of Pasteur (R. L. Devonshire, Trans.). Constable & Company.

Author biography

R. Hans Phaf is an associate professor at the Amsterdam Brain and Cognition Center (ABC) of the University of Amsterdam. He combines experimental research with connectionist and evolutionary modelling in the fields of attention, memory, emotion, and contrasts between conscious and non-conscious processes. His work has been published in journals such as Emotion; Journal of Cognitive Neuroscience; Cognitive, Affective, & Behavioral Neuroscience.