


SYMPOSIA PAPER

The Psychologist's Green Thumb

Sophia Crüwell 

Department of History and Philosophy of Science, University of Cambridge, Cambridge, UK
Email: slbc2@cam.ac.uk

(Received 19 April 2023; revised 01 August 2023; accepted 13 August 2023; first published online 08 September 2023)

Abstract

The “psychologist’s green thumb” refers to the argument that an experimenter needs an indeterminate set of skills to successfully replicate an effect. This argument is sometimes invoked by psychological researchers to explain away failures of independent replication attempts of their work. In this article, I assess the psychologist’s green thumb as a candidate explanation for individual replication failure and argue that it is potentially costly for psychology as a field. I also present other, more likely reasons for these replication failures. I conclude that appealing to a psychologist’s green thumb is not a convincing explanation for replication failure.

1. Introduction

The idiom “green thumb” comes from gardening, where someone often working with terracotta pots might get green thumbs from the algae growing on the pots. It implies that someone has a particular talent for and success in gardening, and that this talent involves inarticulable properties. In this article, I investigate whether the existence of a skill analogous to the gardener’s green thumb is possible and convincing in psychological research, specifically in the context of failed replications in the replication crisis. I begin by introducing the replication crisis and the origins of the green thumb argument, including two example cases. I then describe and assess possible versions of the green thumb argument. Building on this, I consider the kinds of skill appealed to in green thumb arguments and argue that what is implied in such arguments is incompatible with a range of scientific principles and norms. Finally, I contend that we should be suspicious of appeals to the psychologist’s green thumb, particularly given more likely reasons for individual replication failure. I conclude that appeals to the green thumb argument are not convincing as explanations for failed replications in the context of the replication crisis in psychology.

The replication crisis in psychology began around 2011 and accelerated when several replication projects resulted in a high frequency of unsuccessful replications, which was surprising and concerning to the field. Compared to the original studies, the replications that were carried out in these projects were generally more highly

powered to detect relevant effect sizes and more rigorously carried out, including the preregistration of the experiments in question. Mostly, these were so-called “direct” replications, that is, they tried to replicate the original experiment as precisely as possible. Prior to the replication crisis, this type of research was uncommon, at least explicitly labelled as such in *published* psychology (Makel et al. 2012), and there is some indication that it is still uncommon now (Hardwicke et al. 2022). It is not surprising that there was and is uncertainty and contention in the research community regarding how to approach the results of replication studies. This article aims to critically examine one strategy used by original authors to explain replication failures of their work.

Over the years, many causes for individual and fieldwide replication failure have been proposed, including cognitive biases generally (Bishop 2019; Machery 2021), questionable research practices (Simmons et al. 2011), publication bias (Nelson 2020), an unhelpful incentive system (Heesen 2018), low prior probability of hypotheses (Bird 2021), and unduly high confidence in the results of psychological research (Crüwell 2023). These explanations are important but are not the focus of this article. A further explanation, particularly for individual replication failure, is a lack of experimenter skill on the part of the experimenter (Strack 2017). This last candidate explanation, which I call the “psychologist’s green thumb” argument, is what I assess in this article.

2. Origins and examples of the psychologist’s green thumb

The psychologist’s green thumb argument is primarily brought forward by the original authors of studies that failed to replicate. Replicators have been called “novice researchers” (Cunningham & Baumeister 2016, 1) who are “incompetent or ill-informed” (Bargh 2012), they have been accused of “profound naiveté” (Mitchell 2014), and of lacking the “flair, intuition, and related skills” (Baumeister 2016, 156) required for successful replication. To further illustrate how the psychologist’s green thumb argument is invoked, I would like to discuss two examples: the cases of ego depletion and facial feedback.

Ego depletion is a purported phenomenon whereby self-control is a depletable resource. An example of this would be that you are more likely to eat a cookie if you have worked hard and have depleted your self-control. In the study in question, participants were either given a tedious task or an easy task and were subsequently given the same harder task—the idea being that those whose egos have been depleted would perform worse. The original study by Baumeister et al. (1998) was carried out in one lab, testing 67 people, and found what is conventionally thought to be a very large effect size. This original study has been cited more than 7,000 times. There have been two large-scale replication studies involving 23 and 36 labs, respectively, which tested 2,141 and 3,531 participants, respectively (Hagger et al. 2016; Vohs et al. 2021). The replication studies were more highly powered to detect a more likely effect size, preregistered, and transparently reported. Both replications resulted in nonsignificant results with small effect sizes. The original author’s response to these failed replications included the following: “Getting a significant result with $n = 10$ often required having an intuitive flair . . . Flair, intuition, and related skills matter much less with $n = 50$ ”; “a broadly incompetent experimenter can amass a series of impressive publications simply by failing to replicate other work” (Baumeister 2016,

156); Baumeister also described replicators as “novice investigators” (Cunningham & Baumeister 2016, 4).

Another example of the green thumb argument being used can be seen in the literature on the facial feedback hypothesis (Strack et al. 1988; Wagenmakers et al. 2016). The facial feedback hypothesis is based on ideas from embodied cognition, specifically that your facial expression or muscles affect how you feel. In this study, participants held a pen in their mouth—either between their teeth, activating the same muscles as when smiling, or between their lips, activating the same muscles as when pouting (Strack et al. 1988; see Figure 1). This appeared to affect how funny participants found cartoons they were given. This original study was carried out in a single lab, testing 92 participants, and found a significant mean difference. It was an influential study and has been cited more than 2,500 times. The replication study by Wagenmakers et al. (2016) was run by 17 labs, which tested a total of 2,124 participants and found a nonsignificant result. The replication research was highly powered to detect a relevant effect size and was preregistered and transparently reported, including shared data, materials, and code. Nevertheless, Strack has been dismissive of these results, and has claimed a bias toward negative replication results, arguing that “it is easier to be successful at non-replications while it takes expertise and diligence to generate a new result in a reliable fashion” (Strack 2017, 3).

The features that are appealed to by original authors in such cases include diligence, expertise, experience, flair, and intuition. In some cases, a lack of these features can be ruled out as a compelling explanation of replication failure, as they are clearly present in the replication projects in question. In the examples considered in the preceding text, the replicators seem to exceed the original authors in diligence: They collected much larger samples and made their research as transparent as possible. Regarding the features expertise and experience, the replication projects considered previously appear to meet relevant standards. The first author of Hagger et al. (2016) is a professor in social psychology and had previously published meta-analyses on the topic of ego-depletion. The larger replication study of this effect, Vohs et al. (2021), was led by one of Baumeister’s collaborators, who was presumably chosen to lead this project because of her relevant expertise and experience. Similarly for facial feedback—the first author of Wagenmakers et al. (2016) is a professor in psychological methods, and while his research focus is not on social psychology in particular, he had prior experience in social psychological research and thus arguably the necessary expertise and experience to carry out a replication study. The remaining candidate features mentioned in the context of appeals to a green thumb are flair and intuition, which seem to describe a certain style and propensity for finding the right kind of results. These features are more vague and less easy to judge. What I call the “psychologist’s green thumb” in the following sections is an abstraction of these features.

The psychologist’s green thumb thus broadly construed seems ad hoc, particularly in the context of replication studies that are, overall, more highly powered, more rigorously carried out, and more transparently reported than the corresponding original studies. At worst, these claims appear to have been made in bad faith, potentially as a defensive and reflexive reaction to a previously uncommonly published type of research. It is interesting to note that there is initial evidence that scientists fear the consequences of a failed replication more than they should, which may point to why some researchers appeal to the psychologist’s green thumb when

other explanations are more likely: Fetterman and Sassenberg (2015) found that scientists overestimated the negative consequences to their career of failed replications of their work, and that it may in fact be reputationally advantageous for original authors whose work failed to replicate to publicly adjust their beliefs. Accordingly, appeals to a green thumb to explain away replication failure have been rejected by the reform movement and by those conducting replication research. In what follows, I present and assess the *best possible case* for this argument.¹

3. The psychologist's green thumb argument

Having described the origins of the psychologist's green thumb and its abstract features, I now outline the psychologist's green thumb as an argument, presenting a weaker and a stronger formulation. I reject the weaker formulation and take the stronger version forward. An initial formulation of the green thumb argument might be achieved in terms of experimental unshared *propositional knowledge*:

A1 The psychologist's green thumb as propositional knowledge

1. The psychologist's green thumb is experimental knowledge that is not shared.
2. This experimental knowledge is needed to make psychological experiments work reliably.
3. Therefore, if the replicators lacked the psychologist's green thumb, they could not have been able to carry out the replication of the experiment successfully.

While this version of the argument is valid, this interpretation of the green thumb as unshared or unarticulated information puts the responsibility for replication failure on the original authors who failed to supply all relevant information. Had the original authors shared all relevant information when describing their study, the replicators would have been able to replicate the experiment faithfully—whether or not this replication would have been successful is a separate question. This formulation of the argument is thus not an accurate representation of the psychologist's green thumb argument as commonly used by original authors.

A further, stronger formulation of the green thumb argument understands the green thumb to be a form of tacit knowledge or skill. Here, the psychologist's green thumb is experimental skill that cannot or cannot easily be shared:

A2 The psychologist's green thumb as skill

1. The psychologist's green thumb is experimental knowledge or skill that cannot easily be shared.
2. This experimental knowledge or skill is needed to make psychological experiments work reliably.
3. Therefore, if the replicators lacked the psychologist's green thumb, they could not have been able to carry out the replication of the experiment successfully.

¹ I want to note here that there are many unambiguous and uncomplicated cases of creative skill in science that do not fall under possible green thumb explanations. These are singular incidents that do not need to or cannot be replicated. They include creativity in hypothesis generation, original experimental designs, and novel inference.

If the green thumb is experimental skill that takes the form of tacit knowledge or skill that, at least seemingly, cannot or cannot easily be shared, and if that knowledge is needed to be a “successful” experimenter, then a replicator may appear to lack the psychologist’s green thumb if they fail to replicate a result. This is the version of the argument that promises to be more convincing and is thus used hereafter. In the next section, I consider whether this argument is convincing in the context of replication failures in psychology.

4. Is psychology soup, pastry, or potion?

Analogously to cooking or gardening, the green thumb as skill argument can superficially be at least somewhat persuasive: There is a clear difference between apple sauce made by me and that made by my grandmother, or by a Michelin star chef. And while that difference is partly due to practice, it seems plausible that it is also due to some kind of skill. But what kind of skill is the green thumb? I argue that this depends on how we conceive of (social) psychology. To stay with the cooking analogy, the question is whether research in psychology is like cooking a soup, baking a pastry, or making a potion.

A soup is relatively uncomplicated to make, and domain knowledge is arguably not necessary. You can follow a recipe and cut up vegetables, add hot water and seasoning, and wait for a specified time, and you will achieve a meal that looks like the soup in the cookbook. Pastry baking is more complicated as, in this area, a standard recipe might not be able to convey all the relevant details. There are at least two characteristics that make baking more complicated than soup making: There are more external factors that may come into play (such as air pressure or temperature) and there are more decisions that have to be made based on vague perceptual instructions. For example, an instruction may be that you need to take the cake out of the oven when it is golden brown, but not another shade of brown. Overall, in the case of baking, while the information arguably can be articulated, it may not always be straightforward to reduce that information to instructions that can reasonably be followed by anyone. The final possible skill analogy is potion making. Most, if not all, humans cannot make potions, as they lack the crucial, inarticulable, and nontransferable skill of magic.

Is psychology research more like making a soup, is it like pastry baking, or is it like potion making? It seems fair to say that soup is too simplistic a view of psychological research (cf. Collins 1975; the “algorithmical model”), but pastry baking seems to be an accurate analogue. Similarly to pastry baking, expertise and diligence are clearly important in psychological research. While there are many details to learn to consider, and not all these details are easily described, it is possible to learn how to bake a pastry: The skills needed in pastry making are intersubjectively transferrable, both between subjects and between related but nonidentical contexts. With proper instruction by a competent croissant-maker, someone can learn how to make a croissant (cf. Collins 1975; the “enculturational model”). Similarly, someone who has never made a croissant but can make a variety of different but related pastries (brioches, pies, etc.) will be able to bake a croissant from basic instructions. If the psychologist’s green thumb amounts to the skill involved in pastry baking, then this seems compelling: It makes sense that not just anyone can replicate a study, as at least

some domain knowledge and practice in relevant experimental skills are required for successful experimentation.

Thus, a charitable understanding of the green thumb argument is that the original authors see the replicators as having attempted to replicate their experiment as if it was a soup although it required specialized pastry baking skills, which were difficult to fully articulate in the methods section of the original article. However, if we insert such an enculturated, pastry-baking skill into the green thumb argument, is this argument persuasive to the extent that it explains replications that led to nonsignificant results and very small effect sizes, rather than just less clearly successful replications? Construing skill broadly, it has been found not to be related to whether or not a replication was successful in the 100 studies included in the Open Science Collaboration psychology replication project (using the proxy measure of *h*-index: Protzko & Schooler 2020; using the proxy measure of number of publications: Bench et al. 2017). Using a more specific measure of skill, in the examples described in section 2, I argued that the replication teams were both diligent and had field-relevant narrow expertise or skill, meaning that they were arguably appropriately enculturated—they were also pastry chefs (some even sous-chefs in the original authors' kitchens), and they were not attempting the experiments in a souplike fashion. The researchers involved might not have previously conducted the experiment at issue, but they did acquire all relevant skills from nonidentical but sufficiently related contexts. Given that the researchers in the preceding examples were also pastry chefs, appeals to a pastry-baking green thumb cannot explain the drop from a significant finding and large effect size for ego depletion to a small and nonsignificant effect size. Thus, if these psychology researchers were unsuccessful in their replication attempts due to a lack of green thumb, then the kind of green thumb that would be needed for successful replication here seems more similar to the skill required to make a potion than pastry. Psychology research is like baking pastry, but the green thumb appealed to by some original authors to explain replication failures assumes potion-making skills. I will further examine this discrepancy in the following section: Is a potion-like green thumb a—or at least the most—convincing explanation for individual replication failure?

5. Against the psychologist's green thumb as an explanation for replication failure

In what follows, I argue that appeals to a potion-like green thumb are unconvincing and costly for psychology and that there are more convincing and less costly explanations for purported green thumb replication failures.

5.1 Appeals to a potion-like green thumb are costly

I will now consider the costs to psychology of appealing to a potion-like green thumb to explain replication failure. To this end, I examine whether it is compatible with different sets of scientific principles, specifically the conditions for objectivity as proposed by Helen Longino (1990) and the epistemic norms proposed by Robert Merton (1942). I argue that a potion-like green thumb is not compatible with either account and is thus not a convincing explanation for replication failure.

Consider the conditions for objectivity as proposed by Helen Longino (1990):

1. Recognized avenues for criticism;
2. Shared standards;
3. Community response; and
4. Equality of intellectual authority.

First, criticizing research is an important part of science. Appealing to a green thumb indirectly affects this condition, as any criticism or challenge of a researcher's work using replication can be suppressed in that manner. The second condition—shared standards for evaluating research and accepting or rejecting theories—is violated by a potion-like green thumb, as it is a separate and inaccessible standard. Third is the importance of community response, such as in the form of belief change based on criticism. This condition is not met if appeals to a potion-like green thumb are allowed as it obviates the need for a belief change when faced with conflicting evidence. Indeed, this has been found in the literatures on facial feedback and ego depletion, which have not sufficiently taken unsuccessful replications into account (Hardwicke et al. 2022). The final condition concerns shared and equal intellectual authority, which makes it possible to challenge and criticise. Again, appeals to a potion-like green thumb violate this condition, as its existence prevents intersubjective consensus. An original author with a potion-like green thumb necessarily has authority over the replicator without such a green thumb. In fact, the green thumb may be at least partly responsible for a historical lack of published replications. Early career researchers often assumed that they were responsible for any replication failure and did not (try to) publish their failed replications (Lubega et al. 2022). Overall, for those wanting to appeal to such a green thumb argument, the price is violation of all conditions for objectivity on Longino's account.

In fact, a potion-like green thumb is costly due to its incompatibility with a *range* of scientific principles. Consider also the epistemic norms proposed by Robert Merton (1942):

1. Communism/communality;
2. Universalism;
3. Disinterestedness; and
4. Organized skepticism.

Again, it is straightforward to see how appealing to a potion-like green thumb violates these norms as it makes intersubjective consensus virtually impossible. The first norm is communism or communality: The idea that scientific knowledge is public and owned communally. A potion-like green thumb is incompatible with this, as it is secret, inaccessible knowledge. The second norm is universalism, meaning that scientific validity does not depend on the identity of the researcher. This norm is violated by potion-like green thumb, as it fundamentally depends on who did the research. This norm violation also raises questions of robustness: If the effects in question are so fragile that another (social) psychologist cannot find them, it is not clear whether they can be effects of broader interest. Third is the norm of disinterestedness, which states that the goal of research is to advance scientific

knowledge rather than personal gain. Whether a potion-like green thumb is compatible with this depends on the motivations of the researchers appealing to it, and whether it is possible to distinguish if they are defending “their” result to save face or for scientific reasons. The final norm is organized skepticism, which similarly to Longino’s conditions for objectivity holds criticism to be crucial to science. Appeals to a potion-like green thumb at least indirectly hinder this norm: The existence of conflicting evidence may be easily ignored by appealing to such a green thumb. On the whole, appealing to a potion-like green thumb to explain replication failure seems incompatible with established scientific principles, here exemplified by the conditions for objectivity proposed by Longino (1990) and the epistemic norms proposed by Merton (1942).

5.2 Alternative explanations for green thumb replication failures

Such potential violations of scientific norms and principles should make us suspicious of appeals to a potion-like green thumb in psychological experiments in the context of replication failures. On top of this, I now argue that there are other, more likely explanations for alleged green thumb replication failures.

One alternative explanation could be that the replicators did not accurately perform the replication. This could happen in at least three ways (Luttrell et al. 2017): The replication might not be exact enough, have included conceptually necessary adaptations, or have accounted for theoretically relevant moderators. Importantly, this explanation is empirically testable, and should be empirically tested when invoked (see e.g., *ibid.*, who tested this explanation in one case and found that the replication was indeed not carried out appropriately).

Such a replication failure may also happen because the effect under investigation is weak or nonrobust. This seems to be a reasonable explanation, particularly for highly context-sensitive effects, as is common in social psychology. A further possible explanation is that the original authors used research practices in the original study that increased the false-positive rate (Simmons et al. 2011).

Another possible explanation is that the original authors failed to articulate important background information or assumptions or wrote poor methods sections. In fact, it has been argued that failed replications may result in the explication of hidden auxiliary hypotheses representing tacit knowledge or skill, leading to *productive* advances through “operational analysis” (Feest 2016). The appropriate articulation of important information and assumptions, such that the experiment can be replicated by researchers in the same area, is the responsibility of the original authors.

A final potential explanation is the original authors’ unduly high confidence in the results of standard psychological research. Replication failures in social psychology should not be surprising, either due to low prior probability of hypotheses in psychology (Bird 2021) or due to a broader crisis of inference (Crüwell 2023). According to the latter approach, the original authors’ posteriors are unduly high following their own studies, given what we know about how psychology research was and is done. In this framework, appeals to the green thumb can thus be seen as an attempt to explain the discrepancy between the original authors’ unduly high

posteriors and the conflicting new evidence. If the original authors adjusted their inferences and somewhat decreased their confidence in their original results, they would not be as surprised about conflicting replication results.

Overall, there are a range of explanations that seem to be better supported and more parsimonious than that of a potion-like green thumb, particularly given the high cost of appeals to such a green thumb discussed in the previous subsection.

6. Conclusion

Appeals to the psychologist's green thumb are not convincing explanations of replication failures, both because appealing to a green thumb that impacts replication results is costly for psychology and the researchers making use of this argument, and because there are many more likely explanations for individual replication failure. This view leaves space for creativity and green thumb skills in appropriate contexts such as hypothesis generation.

Acknowledgments. I am grateful for helpful comments from Jacob Stegenga, Edouard Machery, and Alexander Bird, as well as for discussion with Ina Jäntgen, Cristian Larroulet Philippi, Benjamin Chin-Yee, Hermann Crüwell, Michaela Egli, Arthur Harris, Oliver Holdsworth, Adrià Segarra, Charlotte Zimmel, and the audience at the 2022 Philosophy of Science Association Biennial Meeting.

References

- Bargh, John A. 2012. "Nothing in Their Heads." [Deleted blogpost, archived at <https://replicationindex.com/wp-content/uploads/2020/07/bargh-nothingintheirheads.pdf>]
- Baumeister, Roy F. 2016. "Charting the Future of Social Psychology on Stormy Seas: Winners, Losers, and Recommendations." *Journal of Experimental Social Psychology* 66:153–58. <https://doi.org/10.1016/j.jesp.2016.02.003>
- Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, & Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74 (5):1252–65. <https://doi.org/10.1037//0022-3514.74.5.1252>
- Bench, Shane W., Grace N. Rivera, Rebecca J. Schlegel, Joshua A. Hicks, & Heather C. Lench. 2017. "Does Expertise Matter in Replication? An Examination of the Reproducibility Project: Psychology." *Journal of Experimental Social Psychology* 68:181–84. <https://doi.org/10.1016/j.jesp.2016.07.003>
- Bird, Alexander. 2021. "Understanding the Replication Crisis as a Base Rate Fallacy." *The British Journal for the Philosophy of Science*, 72:965–93. <https://doi.org/10.1093/bjps/axy051>
- Bishop, Dorothy. 2019. "Fixing the Replication Crisis: The Need to Understand Human Psychology." *APS Observer* 32 (10). <https://www.psychologicalscience.org/observer/fixing-the-replication-crisis-the-need-to-understand-human-psychology>
- Collins, H. M. 1975. "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics." *Sociology* 9 (2):205–24. <https://doi.org/10.1177/003803857500900202>
- Crüwell, Sophia. 2023. "Reframing the 'Replication Crisis' as a Crisis of Inference." [Unpublished manuscript].
- Cunningham, Michael R., & Roy F. Baumeister. 2016. "How to Make Nothing Out of Something: Analyses of the Impact of Study Sampling and Statistical Interpretation in Misleading Meta-analytic Conclusions." *Frontiers in Psychology*, 7:1639. <https://doi.org/10.3389/fpsyg.2016.01639>
- Feest, Uljana. 2016. "The Experimenters' Regress Reconsidered: Replication, Tacit Knowledge, and the Dynamics of Knowledge Generation." *Studies in History and Philosophy of Science Part A* 58:34–45. <https://doi.org/10.1016/j.shpsa.2016.04.003>
- Fetterman, Adam K. & Kai Sassenberg. 2015. "The Reputational Consequences of Failed Replications and Wrongness Admission among Scientists." *PLOS ONE* 10(12): e0143723. <https://doi.org/10.1371/journal.pone.0143723>

- Hagger, M. S., N. L. Chatzisarantis, H. Alberts, C. O. Anggono, C. Batailler, A. R. Birt, R. Brand, et al. 2016. "A Multilab Preregistered replication of the Ego-Depletion Effect." *Perspectives on Psychological Science* 11 (4):546–73. <https://doi.org/10.1177/1745691616652873>
- Hardwicke, Tom E., Robert T. Thibault, Jessica E. Kosie, Joshua D. Wallach, Mallory C. Kidwell, & John P. A. Ioannidis. 2022. "Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017)." *Perspectives on Psychological Science* 17 (1):239–51. <https://doi.org/10.1177/1745691620979806>
- Heesen, Remco. 2018. "Why the Reward Structure of Science Makes Reproducibility Problems Inevitable." *The Journal of Philosophy* 115 (12):661–74. <https://doi.org/10.5840/jphil20181151239>
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Lubega, Nasser, Abigail Anderson, & Nicole C. Nelson. 2022. *Experience of Irreproducibility as a Risk Factor for Poor Mental Health in Biomedical Science Doctoral Students: A Survey and Interview-Based Study* [Preprint]. MetaArXiv. <https://doi.org/10.31222/osf.io/h37kw>
- Luttrell, Andrew, Richard E. Petty, & Mengran Xu. 2017. "Replicating and Fixing Failed Replications: The Case of Need for Cognition and Argument Quality." *Journal of Experimental Social Psychology* 69:178–83. <https://doi.org/10.1016/j.jesp.2016.09.006>
- Machery, Edouard. 2021. "A Mistaken Confidence in Data." *European Journal for Philosophy of Science* 11: 34. <https://doi.org/10.1007/s13194-021-00354-9>
- Makel, Matthew C., Jonathan A. Plucker, & Boyd Hegarty. 2012. "Replications in Psychology Research: How Often Do They Really Occur?" *Perspectives on Psychological Science* 7 (6):537–42. <https://doi.org/10.1177/1745691612460688>
- Merton, Robert K. 1942. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*, 267–78. Chicago: University of Chicago Press.
- Mitchell, Jason. 2014. "On the Emptiness of Failed Replications." [Deleted essay, archived at https://web.archive.org/web/20140708164605/http://wjh.harvard.edu/~jmitchel/writing/failed_science.htm]
- Nelson, Nicole. 2020. "Towards an Expanded Conception of Publication Bias." *Journal of Trial and Error* 1 (1):52–58. <https://doi.org/10.36850/mr2>
- Protzko, John, & Jonathan W. Schooler. 2020. "No Relationship between Researcher Impact and Replication Effect: An Analysis of Five Studies with 100 Replications." *PeerJ* 8:e8014. <https://doi.org/10.7717/peerj.8014>
- Simmons, Joseph P., Leif D. Nelson, & Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11):1359–66. <https://doi.org/10.1177/0956797611417632>
- Strack, Fritz. 2017. "From Data to Truth in Psychological Science. A Personal Perspective." *Frontiers in Psychology* 8:702. <https://www.frontiersin.org/article/10.3389/fpsyg.2017.00702>
- Strack, Fritz, Leonard L. Martin, & Sabine Stepper. 1988. "Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis." *Journal of Personality and Social Psychology* 54 (5):768. <https://doi.org/10.1037/0022-3514.54.5.768>
- Vohs, Kathleen, Brandon Schmeichel, Sophie Lohmann, Quentin Gronau, Anna Finley, Sarah Ainsworth, Jessica Alquist, et al. 2021. "A Multi-Site Preregistered Paradigmatic Test of the Ego Depletion Effect." *Psychological Science* 32 (10):1566–1581. <https://doi.org/10.1177/0956797621989733>
- Wagenmakers, E. J., T. Beek, L. Dijkhoff, Q. F. Gronau, A. Acosta, R. B. Adams Jr, D. N. Albohn, et al. 2016. "Registered Replication Report: Strack, Martin, & Stepper (1988)." *Perspectives on Psychological Science* 11 (6):917–28. <https://doi.org/10.1177/1745691616674458>