This is the reason for external replication: Response to Aungle et al. (2026)

Nicholas J. L. Brown[1]
Andrew Gelman[2]

[1]Department of Psychology, Linnaeus University, Växjö, Sweden, and [2]Department of Statistics and Department of Political Science, Columbia University, New York.

24 Mar 2026

Abstract. In an earlier article we addressed a controversy regarding a form of mind-body healing, arguing that a recent paper had overstated evidence from experiments and from literature review. In reaction, one of the authors of that paper disputed our claims. Here we explain why we remain skeptical.

Overview

Aungle and Langer (2023) argued in favor of a form of mind-body healing, with the evidence being a new study of the relation between observed healing of minor injuries and the perceived passing of time, along with a literature review of previous experiments finding similar effects. In Gelman and Brown (2024) we expressed skepticism, pointing to problems with the analysis and inference from Aungle and Langer's experimental data and to ways in which their citations of past literature were inaccurate and misleading. In response, Aungle et al. (2026) replied to our criticism, arguing that their statistical evidence remains strong after correcting for their data analysis error and also arguing that we were sloppy in our description of the literature they had cited.

Our reason for writing the 2024 paper was not primarily to address mind-body healing but rather to consider ways in which researchers can fool themselves and others, first by overestimating the strength of evidence from their own studies and second by mistakenly finding confirmation from the literature. To the extent that ideas of mind-body healing can be discovered, isolated, and made useful, we believe this would be best done by experiments with clearly-defined treatments, careful measurements, and validated by external replication.

We have been offered the opportunity by this journal to respond to Aungle et al. (2026), which we do here in as brief a form as we feel is appropriate at this stage of the discussion. Our 2024 paper makes two key points that we still stand by:

1. We do not find the data from the Aungle and Langer (2023) experiment to offer convincing evidence of mind-body healing, nor do we find the theory persuasive—no mechanism beyond a vague reference to placebo effects has been proposed for how the manipulation of time would speed the recovery from a bruise—and this leads us to the question of past evidence.

2. We have not attempted an overview of this entire literature or even of all the papers cited by Aungle and Langer (2023). We looked carefully at three of the citations that were offered as support for their thesis, and none were as described by Aungle and Langer (2023).

Aungle et al. (2026) is well written and can be read as making a strong case for the presence of strong mind-body healing effects in general, in the experiment of Aungle and Langer (2023), and in the literature cited therein. However, the main thing this demonstrates to us is the malleability of language, in that Aungle et al. were able to write a persuasive argument even in the absence of any good evidence. Verbal arguments are valuable—if we did not think so, we would not have written our 2024 paper or the present rejoinder—but when it comes to assessment of scientific theory there is no substitute for external replication.

Specific responses

Aungle et al. (2026) offered two major disagreements with our earlier paper. First, they argued that their experimental results (after correction, the *t* scores are 3.0 for the 56 min vs. 14 min condition and 0.9 for the 28 min vs. 14 min condition) provide strong evidence in favor of their hypothesis. Second, they argued that our discussion of the literature on mind-body healing was flawed because we discussed only three out of the many papers cited by Aungle and Langer (2023).

We will discuss these in turn.

First we need to explain why we do not consider the experimental evidence from Aungle and Langer (2023) to be convincing even though we agree that, in some contexts, *t* scores of 3.0, and even 0.9, can provide reasonable evidence. The problems here are well known

from reports of the replication crisis in science: uncontrolled researcher degrees of freedom and theoretical challenges of interpreting departures from the null as support for a particular explanation (in this case, mind-body healing rather than measurement bias or, for example, as an interaction between psychological factors and the purported mechanism of action of cupping itself). This is not a problem that can be resolved using Bayesian inference or any other statistical paradigm, as the fundamental problem is a lack of strong evidence, which could be seen in classical terms via adjustment for multiple potential comparisons or a multiverse analysis, or in Bayesian terms with a hierarchical model, which in this case would find no evidence for the presence of large or consistent effects.

That said, Aungle et al. (2026) did point out a mistake in our interpretation of the experiment in question, as we remarked that the participants in the experiment had performed different computer activities under different conditions, without realizing that those different activities had been performed *after* the outcome measurements and thus had no effect on the results. Although this does not alter our conclusion (as this was just one of many potential alternative explanations for their findings), we apologize and acknowledge our error, and we respect the challenge of running experiments in the real world. Indeed, it is in large part from our respect for the work of experimental psychologists and the importance of their project that we wrote our critique, because it saddens and frustrates us to see this sort of care and effort devoted to construction and overinterpretation of noisy datasets. To the extent that researchers care about mind-body healing, we hope they can move away from black-box experimentation, no matter how cleverly designed and carefully executed, and toward more direct investigation of potential mechanisms.

To return to the statistical issue: Aungle et al. (2026, p. 3) wrote that we “suggested that Aungle and Langer may have tested a number of factors that influenced healing and only reported those that turned out significant.” But we did not intend to make any such suggestion. Rather, we were merely noting that in an experiment with many forking paths in data coding and analysis, it is usual for the steps taken in a non-preregistered analysis to depend on data, and in this setting there would be no difficulty in coming up with explanations for just about any pattern in data. As discussed by Gelman and Loken (2014), this problem arises without any assumption that researchers tried out multiple analyses on the data at hand; it is enough that coding and analysis decisions are made after looking at the data. This is not to say that no such studies can be trusted or that nothing can be learned from non-preregistered analysis, but rather that it is a mistake to naively identify a p value less than 0.05 as strong evidence of a positive effect. In addition, for the experiment reported in Aungle and Langer (2023), the estimated slope variance (with a standard deviation of

about twice the mean in one case and eight times the mean in another) implies that any participant-level effects are extremely uncertain; a stable sign at the level of the overall sample does not guarantee a reliable, replicable effect for a given individual. Such heterogeneity would limit the interpretability and practical significance of any effect.

In the second section of their paper, Aungle et al. (2026) critiqued our review of the mind-body literature, arguing that we should have looked at 28 articles cited by Aungle and Langer (2023) rather than merely three (plus the Aungle and Langer article itself) that we commented on. But our aim was not to assess the general placebo effect (a topic which itself is controversial; see Birguslatro, 2024), but rather to look into the evidence provided by Aungle and Langer; a comprehensive review of all 28 citations would have been beyond the scope of our commentary.

We looked into three of the cited references because they looked interesting; here are the descriptions from Aungle and Langer (2023): (1) “patients who received physician assurances after skin pricks healed significantly faster,” (2) “the suggestion that one had touched poison ivy resulted in stronger symptoms than actually touching poison ivy,” and (3) “If a person who does not exercise weighed themselves, checked their blood pressure, took careful body measurements, wrote everything down, maintained their same diet and level of physical activity, and then repeated the same measures a month later, few would expect exercise-like improvements. But in a study involving hotel housekeepers, that is effectively what the researchers- found.” If any of these claims were true—let alone all three—this could represent strong evidence of relevant effects. On the other hand, if these three past studies did not go as described, this casts into question the idea that the time-healing experiment could be expected to work as advertised.

Unfortunately, as we explained in detail in our commentary (Gelman & Brown, 2024), none of the above three descriptions of past studies were accurate. (1) The skin-prick experiment was subject to researcher degrees of freedom so that the appearance of a statistically-significant ($z = 1.96$, $p = 0.05$) result could easily have arisen just by chance (as discussed in general terms by Simmons et al., 2011) or by alternative causal pathways having nothing to do with mind-body healing. (2) The poison-ivy experiment was a reference to Ikemi and Nagakawa (1962), who included a qualitative study of 13 boys who were exposed on one arm to the poisonous leaves of the lacquer or wax tree (not actually poison ivy) and on the other arm to inert leaves, but were told that the exposures were the reverse. This was an uncontrolled experiment which, according to a review article from 2015, had never been replicated. (3) The exercise study did not actually find that the housekeepers “maintained

their same diet and level of physical activity." Self-reported exercise actually increased in this study, but even if we were to accept the position of Aungle et al. (2026) that those reports merely report *perceived* exercise, the study offers no evidence that there were no actual changes in diet and exercise during the study period.

The problem in all three of these cases is that Aungle and Langer (2023) made strong claims based on inaccurate descriptions of past studies, and this calls into question the notion that their research hypothesis is well supported by the literature. In their reply, Aungle et al. (2026) cited an article by Zahrt and Crum (2017)—which was not one of the 28 articles cited by Aungle and Langer—that relied on secondary observational data with considerable analytic flexibility (e.g., in the selection of covariates from among the dozens of variables in the NHANES dataset). Again, we recommend external preregistered replication—if it is really true, for example, that a one-sentence reassurance could reliably reduce short-term pain, this would be sensational news, and it would make sense to study this definitively rather than relying on noisy uncontrolled studies performed by advocates of the methods in question.

In the same section of their reply, Aungle et al. (2026, p. 5) wrote, "We question the basis of Gelman and Brown's claim that the findings of Crum and Langer (2007) are unreplicable. . . . No exact replication has been reported, so we do not know." We do not dispute Aungle et al.'s observation that we cannot know if a study will replicate or not in the absence of any sort of replication attempt. But that absence of a replication attempt does not make that study immune to criticism of its methods and theoretical basis in the meantime. The promoter of a claim, particularly one whose proposed theoretical mechanism remains underspecified other than the general invocation of the placebo effect, should not expect their conclusions to be considered as scientific fact in the absence of external replication. Acting as if the evidence from a study is stronger than it is, and presenting the literature as more conclusive than it is, leads to a culture of what we might call "drive-by studies," whereby researchers can spray theoretically implausible claims around on the basis of results derived from noisy data and forking paths and then move on, without any serious consideration of alternative explanations. Such a course of action is hardly discouraged by the current academic publishing system, which overwhelmingly favors the production of novel results over the careful refinement of existing claims (Nosek et al., 2012).

In their final section, Aungle et al. (2026) criticized our assertion that Aungle et al.'s (2023) claims of the healing effects of manipulating perceived time have no clear theoretical justification, stating that "The idea that cognitive processes can impact physiological outcomes is supported by decades of rigorous, well-replicated studies in psychology and

medicine.” However, this does not do anything to establish the specific claim at issue. There is no reason to believe that there exists a universal placebo effect that can be harnessed to produce whatever (positive) effect we might like it to, such as accelerating tissue repair. Several millennia of human observation support the idea that birds can fly, but this does not provide evidence for a claim that pigs can also do so. We stand by our assessment that the studies that we critiqued, principally Aungle and Langer (2023), do not provide any sort of coherent theoretical framework for the specific claims that they make regarding aspects of a purported phenomenon of mind-body unity.

The larger picture

We can also place this discussion in the context of public discourse. The work by Langer and her collaborators on mind-body healing has been widely publicized in the news and social media, and there has also been a long history of criticisms of this work. In our 2024 paper we focused on a single study and then traced back some of its references to address the interesting puzzle of how this seemingly strong published literature could exist alongside a lack of accepted theory and a steady stream of criticism. We concluded that Aungle and Langer (2023) were able to attain a convincing appearance of evidence by a combination of overstatement of evidence from their study, not accounting for forking paths and potential alternative explanations in their analysis, and misrepresenting the literature, including their own previous work. We are not saying that any of this was done intentionally; in science it is easy to drift in the direction of one's expectations. Indeed, in their rebuttal to us, Aungle et al. (2026) argued that it is we who have been biased by our preconceptions. This sort of thing is always possible and can happen to anyone. We are not claiming to be certain that there are no effects; we are merely saying that the data provided by Aungle and Langer (2023) do not provide good evidence for their claims, and neither do their references, when read carefully.

The claims of Aungle and Langer (2023), as well as those in many of the articles that they cited are, as the saying goes, “big if true” (Gelman, 2022). The claimed effects would have extraordinary implications for our understanding of human physiology, medical practice, and public health policy. Speaking for ourselves as researchers, if we really imagined that we had discovered a mechanism as powerful as was purportedly found in these studies, we would be looking for funding for large-scale replication studies, with a view to moving on to clinical trials and widespread adoption as fast as possible, and we would start by trying to lock down some of the results by organizing external preregistered replications. We argue for replications not just to provide convincing evidence of a nonzero effect but to understand how the effect, to the extent that it exists, would be relevant in the real world.

This brings us to another salient feature of the mind-body literature: If a reliable and reproducible mechanism could be identified, it would be worth billions of dollars, in the form of either profits for therapists or reduced costs for healthcare systems. Yet, no such developments ever seem to take place. It reminds us somewhat of extra-sensory perception or dowsing: If people were really able to detect underwater streams, or even oil fields, by walking around until a pair of sticks swung to cross each other, governments and corporations would be lining up to hire the most talented dowsers, learn their secrets, and train armies of people to find these hidden natural resources. The fact that they are not tells us something about the plausibility of the dowsers' claims. No amount of similarly-flawed studies can constitute evidence for a claim that has no clearly specified and empirically established physiological mechanism accounting for the magnitude of the reported effect, and which is at least equally well explained by the operation of chance and the human tendency to detect patterns where none exist.

In our earlier paper we identified two problems with Aungle and Langer (2023): flawed statistical analysis and misinterpretation of the literature. The main purpose of our paper was not to criticize that particular article or even the mind-body literature. Rather, as we discussed in our conclusion to that earlier paper, we wanted to point out that these two problems go together. When there is no consistent treatment effect, flawed analysis can still enable researchers to achieve apparent statistical significance. And when there is no good theory, there is a motivation to find strong support from past research—but if there is no true effect, the only way to find such support is to overestimate the strength of evidence in the literature. Again, we are not suggesting that the authors did anything nefarious here; rather, they have put themselves in the difficult position of seeking strong and consistent evidence for a phenomenon when such evidence may not exist at all.

References

Aungle, P., Chen, D. L., & Holmes, N. P. (2026). Beyond statistical myopia: Replying to a misguided critique of mind-body research. *Advances in Methods and Practices in Psychological Science*, *9*(1), 1–9. https://doi.org/10.1177/25152459261417257

Aungle, P., & Langer, E. J. (2023). Physical healing as a function of perceived time. *Scientific Reports*, *13*, Article 22432. https://doi.org/10.1038/s41598-023-50009-3

Birguslatro (2024, November 13). *A case against the placebo effect*. https://carcinisation.com/2024/11/13/a-case-against-the-placebo-effect/

Gelman, A. (2022). How should scientific journals handle “Big if true” submissions? *Chance*, *35*(2), 41–43. https://doi.org/10.1080/09332480.2022.2066415

Gelman, A., & Brown, N. J. L. (2024). How statistical challenges and misreadings of the literature combine to produce unreplicable science: An example from psychology. *Advances in Methods and Practices in Psychological Science*, *7*(4), 1–13. https://doi.org/10.1177/25152459241276398

Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University. https://sites.stat.columbia.edu/gelman/research/unpublished/forking.pdf

Ikemi, Y., & Nakagawa, S. (1962). A psychosomatic study of contagious dermatitis. *Kyushu Journal of Medical Science*, *13*, 335–350.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

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*(11), 1359–1366. https://doi.org/10.1177/0956797611417632