ERROR REPORT

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nature Human Behaviour, 3*(3), 251-256. https://doi.org/10.1038/s41562-018-0520-3

Reviewer: **R. Chris Fraley,** University of Illinois at Urbana-Champaign
Author response: **Nick Light**, University of Oregon
Recommender: **Jamie Cummins**, University of Bern

DECISION & RECOMMENDATION

Based on the reviewer's report and the author's response, I am returning the decision that the original article contains **Minor Errors** (see below for the ERROR rubric of error report categorisations). That is, errors that have the benefit of being detectable thanks to the presence and sharing of research materials, but whose scope and implications are minor. The detected errors do not rise to the level where I would recommend that a correction be issued. A decision of *minor errors* entails the publication of the report, authors response, and recommendation on the ERROR website (error.reviews). The authors are also asked that, in future, they recognise the minor errors associated with their manuscript in future discussions of the article.

DECISION	RECOMMENDATION
No errors	No additional action beyond publication of the error report and reply
Minor errors	No additional action beyond publication of the error report and reply, plus recommendation that the authors appropriately recognise these errors in future discussions of the article. *Note that we expect most published research to fall into this category.*
Indeterminable	No determination could be made regarding the presence or absence of important potential errors. Recommendation that the authors appropriately recognise this lack of verifiability in future discussions of the article. *Note that this is considered a less desirable outcome than "Minor errors" because in that case the errors are verifiably minor in nature.*
Moderate errors	Correction notice (minor). *Note that numerous 'minor' errors might constitute a moderate error in aggregate.*
Major errors	Correction notice (major) / may warrant an expression of concern.
Severe errors	Retraction

RECOMMENDER'S REPORT

Jamie Cummins

Firstly, I would like to thank both Prof. Fraley (reviewer) and Prof. Light (corresponding on behalf of the authorship team) for their cooperation throughout this review. Prof. Light's response letter begins with highlighting his initial reservations towards participating in the ERROR project, including his worries relating to potential career or reputational damage. We think many authors would likely share similar worries when approached for a project such as this. We are all the more grateful for his participation and indeed we hope his positive experience with this process can serve as an example to future authors who feel a similar initial reluctance in participating.
The substance and style of the reviewer report and authors' response embody what we hope to see more of in academic research: acceptance of the possibility that errors occur; inspection and useful discourse about potential errors that is well-documented and verifiable; and acknowledgement and suitable correction when errors are found.

ERROR recommendations are public documents whose function is to (1) communicate the presence or absence of any errors detected, (2) consider their severity, and (3) provide discussion of how similar errors elsewhere might be prevented or detected. Materials for all error reports can be found at [osf.io/fpw4r](osf.io/fpw4r).

Summary of errors detected & how they could be prevented in future

The original article consisted of four studies investigating the relationship between extremity of opposition to certain topics (primarily genetically-modified foods, but also gene therapy and climate change), objective knowledge on these topics, and perceived knowledge around these topics. We asked Prof. Fraley to review Studies 1, 3, and 4, given that Study 2 was analysed using STATA, which Prof. Fraley was less familiar with, whereas Studies 1, 3, and 4 were analysed using R. Readers should note that Study 2 has therefore not been subjected to the ERROR review process.

As detailed in Prof. Fraley's report, a small number of errors in the original paper were found during the review. The openly-available code associated with the manuscript was not fully computationally reproducible, with portions of code referring to objects which were not created in the script, and (more trivially) some of the R packages not loading correctly. Anyone familiar with attempting to computationally reproduce previous work will be aware that these are extremely common issues not at all idiosyncratic to the work of the original authors. In general these issues can be avoided in future work by (i) having a researcher independently attempt to computationally reproduce the results prior to the submission of the manuscript, and (ii) utilising a version management library, such as groundhog, to specify and preserve the specific package versions which were used for processing and analysis (and to document this clearly). An even more low effort solution

could involve using RMarkdown files for scripts and ensuring that the file knits correctly (as undeclared objects would be caught through this process).

Additionally, the issue of computational reproducibility extended to a small number of values reported in the manuscript, where these values were erroneously reported (e.g., as Prof. Fraley noted, average age was reported in the paper as 36.6 but was actually 39.6 based on the analysis code). This small number of issues amounted to what are likely to have been typos, but can be avoided in principle by (i) having an independent researcher specifically check for such typos prior to manuscript submission, and/or (ii) writing manuscripts using a software which can directly integrate the numeric output of the code into the manuscript text, such as RMarkdown.

Prof. Fraley noted several instances in the paper where inferences about differences between relationships/conditions were made based on divergent results from two separate analyses, rather than a single analysis testing this inference directly. This is a relatively common statistical fallacy, and statistical difference cannot be inferred by the difference between a significant vs. non-significant test (or indeed, even two significant tests; Gelman & Stern, 2006) Although the inferences made by the authors in this case stood up to scrutiny with the more appropriate analyses conducted by Prof. Fraley, this can often not be the case in similar scenarios. To avoid this, authors should generally ensure that the statistical inferences they make regarding differences between relationships or conditions are mapped *directly* onto a statistical inference test, rather than based on divergences between separately conducted tests.

Discussion of individual issues raised

Code functionality and computational reproducibility

The reviewer noted issues with (i) referenced objects in the R code which were not created within the code of the authors, (ii) loading a small number of libraries, some of which were resolved by the reviewer updating his R version, and (iii) copy-paste errors between the code and manuscript. I agree with Prof. Fraley that the first two of these issues amount essentially to "hiccups" and do not substantively affect the content of the manuscript or its conclusions. The third issue affects the content of the manuscript in a small number of cases, specifically in relation to two beta coefficients which missed decimal places in Study 1, and an incorrect reporting of mean age for Study 3. Prof. Light agrees in his reply that this amounts to copy-paste errors on the part of the authorship team. Although these do represent errors in reporting, they do not substantively affect the conclusions of the paper, and therefore represent a minor error.

Lack of direct tests for inferences of difference in relationships/conditions

The reviewer noted that there were a number of instances where the authors made inferences relating to differences in relationships between variables based on different patterns between two analyses, rather than a direct analysis of the difference of these relationships. For instance, the authors state in their abstract that "the relationship between self-assessed and objective knowledge shifts from positive to negative at high levels of opposition" in the domain of GMO foods, but not for climate change, and subsequently compare these domains in their discussion. As Prof. Fraley stated, a more appropriate analysis would directly test the effect of "Condition" in this context to infer that something was the case in one condition and not in the other. Prof. Fraley graciously conducted this analysis (as well as others where similar issues emerged), and found that the substantive conclusions remained to be supported by the more appropriate analyses. Since the conclusions in the manuscript do not change, this may be considered a minor error in terms of its consequences for the article. It is important to note that the analytic approach is erroneous and should not be used to make such inferences. However, on balance and in the context of the conclusions not being changed, we label this a minor error in this context.

Pseudo-causal language

The reviewer highlighted the presence of some pseudo-causal language in the manuscript; namely, making within-person claims based on between-person analyses. I generally agree with Prof. Fraley that there is such language in the manuscript, and that this is not wholly appropriate given the analyses presented. Prof. Light similarly acknowledged and agreed that this may be a potential issue. There are certainly instances of this "pull it and it gets longer" language (as Dr. Fraley described it) throughout the manuscript, such as the final sentence "This suggests that a prerequisite to changing people's views through education may be getting them to first appreciate the gaps in their knowledge". In our view, even such hedged statements can mislead readers into thinking that the paper's results mislead readers because the implicit assumptions behind such a causal inference remain undiscussed. Within ERROR, we are mindful that not all researchers agree on this point, and that statements such as these are generally common within the norms of writing in psychological research. Given all of this, I am inclined to say that this does not quite rise to the level of an error, despite the fact that readers may be misled by this type of language.

Unresolved and unexamined aspects

It is reasonable to expect that ERROR reviews will leave some questions unresolved. It is useful to acknowledge the potential for such issues so that ERROR recommendations do not artificially convey that they represent the final word on issues of error detection and correction for a given article. In this case, readers should once again note that Study 2 of the manuscript did not receive any review given that its analyses were conducted in STATA.

There are also dimensions of the paper which, in my opinion, could potentially be explored in more depth in the future. For instance, there are several analyses which are alluded to in the discussion section of the paper but which were not examined by Prof. Fraley. One of these analyses relates to the interaction of effects with political identification; the authors find that political identification interacted with effects for climate change but not for GM foods, but not directly test this statistical difference (i.e., echoing similar issues Prof. Fralay flagged in terms of the lack of direct tests for inferences). Additionally, the authors mention in the discussion they "re-ran the main analyses from all studies including education level as a control", and that all key findings remained significant. However, the additional analyses Prof. Fraley ran were not tested for their robustness with such a control (and indeed, the computational reproducibility of those re-run analyses were also not inspected). To be explicit, this is not to say that any of these examples contain errors; it is simply to highlight that they were not examined in this ERROR review.

I sincerely thank both Prof. Fraley and Prof. Light again for their efforts during this process.

Jamie Cummins
Recommender for ERЯOR

ЯEVIEW

R. Chris Fraley

I. METHODS, MEASUREMENT, AND DESIGN

1. Design `No errors found`

Are there errors in the conceptual design of the study? E.g., flawed randomisation technique

Time spent: 150 mins minutes/hours

--

Here is a brief overview of my error-checking process:

1. I read the paper and made notes of elements to check. I wanted to check the pre-registration (Study 1) against what was done in the paper. I also wanted to reproduce the key results reported in the paper using the data/code provided by the authors.

2. Downloaded the data from the OSF page

3. Downloaded the R code for Studies 1, 3, and 4. (Study 2 code is for STATA, which I do not own. So, I didn't check Study 2.)

4. Double-checked the recoding of variables in Studies 1, 3, and 4 to ensure that there were no typos or copy-and-paste errors. I also checked that reverse coding was consistent across studies.

5. Ran the analysis code in base R, Version 4.3.0. (And, for code requiring a newer version, I installed and used 4.4.1.)

6. Checked reported numbers in the article against results obtained in R.

7. For any analyses I ran independently of the supplied code, I used comments in the newly saved script file to demarcate it clearly.

Example:
```
#----------------------------------------
# Fraley // Begin Fraley code insert k
#----------------------------------------
```

```
        // new code

#----------------------------------------
# Fraley // End Fraley code insert k
#----------------------------------------
```

8. I checked all the numbers reported in the manuscript against what emerged from running the R script. I checked every number for Study 1, but focused more on the key ideas in the subsequent studies because the manuscript itself focused mostly on S1. I didn't check the bootstrapped CIs because they involve randomness. Nonetheless, in each analysis, the reported intervals were in the ballpark that one would expect for bootstrapped analyses.

Broad Summary

Almost everything checked out. A few minor errors are noted below, along with issues that are not "errors" per se, but just "concerns" that would have come up in the review process, had I been a reviewer. (But a reviewer without special theoretical knowledge in this domain.)

One element that was unclear to me was the purpose of the climate change and genetically modified foods distinction. The inclusion of these two contexts, with random assignment to conditions, implies that this was an explicit factor in the original design (i.e., participants were randomly assigned to consider climate issues or genetically modified food issues). And, the way the data are labeled in the Study 1 data file ("cond" for "condition") is consistent with this interpretation, as is the language "two-between-subjects conditions" in the pre-registration (Question 4).

However, the pre-registration doesn't explicitly state that this factor is a factor per se in the analytic section. In practice, the authors analyze the two contexts separately rather than treating this condition as a condition in the design. The analytic section is vague on how this was supposed to be handled. Based on the registration, I was expecting Condition to be an explicit factor in the analyses. The authors make an "error" that is often made in the behavioral sciences (see Gelman & Stern, 2012, for an explanation) where they conclude that something works in one situation (e.g., genetically modified foods) but not in

another (e.g., climate change) because an association emerges in one case but not in the other. But the difference between those associations was not formally tested. Interactions with Condition would allow for this.

Fraley Code Insert 5 addresses this and the results are consistent with the authors' conclusions. To summarize: This is probably a design choice I would consider suboptimal if I were reviewing this in a results-blind manner. Given the data, however, the choice is inconsequential for the conclusions drawn. This choice probably shouldn't be considered an error per se, but it also wouldn't be considered "best practice" in research methods. It is most common in psychological research to treat condition as a factor in an analysis rather than to analyze the data from two conditions separately.

2. Measurement   No errors found
Are there any measures, techniques, or devices that were incorrectly applied or inappropriate for the specific task described in the paper?
Time spent: 5 mins   minutes/hours

3. Preregistration Consistency   No errors found
Are there substantial deviations from the preregistration, particularly undisclosed ones?
Time spent: 10 mins   minutes/hours

I checked the pre-registration against the analyses and methods that were reported for Study 1. To be frank, I wasn't expecting this to line up well because the article reports several analyses that seemed like the kinds of things that would be done to appease reviewers during the review process (e.g., analyzing the two extremity items separately and as a composite). But, in fact, virtually everything reported in the manuscript for Study 1 was part of the pre-registration too. The analyses reported that were not pre-registered (unless I missed it) were the genetics-only items from the scientific knowledge section.

4. Sampling   Not applicable
Is there an error in the sampling strategy? Is the power analysis reproducible? Does the model used for the power analysis match the model in the substantive analyses? Were separate power analyses conducted for all primary analyses?
Time spent: 5 mins   minutes/hours

5. Other Aspects Related to Methods and Measures `No errors found`

Time spent: 5 mins   minutes/hours

II. DATA, CODE, AND STATISTICAL ANALYSES

1. Code Functionality `Errors found`

Does the provided code run without the need to make any adjustments and without errors?
If not, what steps were needed to get it to run (if it was eventually possible)?

I encountered a few issues when running the R script containing the code. Some of these
are just hiccups, but I note them here for completeness. Please note that I didn't check
Study 2 because I do not have a copy of STATA and the analyses for that study were done
in STATA. This includes Figure 3 which uses a data file not included:
"Sydney_GMOData_NLversion.csv"

- Study 3 Johnson-Neyman analyses. The package "probemod" was not compatible with my
version of R (4.3.0). After I updated R to 4.4.1, the probemod analyses ran well and checked
out.

- There was no package or install code for 'stargazer.' When I manually installed it, the
analyses wouldn't run because the package was incompatible with my version of R (4.3.0).
I updated R to 4.4.1 and was able to load stargazer with no problem. However, I received
the following error:

```
> stargazer::stargazer(s1.GMOsubj.q, s1.GMOscilit.q, s1.CCsubj.q,
s1.CCscilit.q, s2.GMOsubj.q, s2.scilit.q,   s3.GMOsubj.q, s3.GMOscilit.q,
s4.GTsubj.q, s4.GTscilit.q, type = "html", out =
"NHBALLstudiesQuadratics2.htm")
Error in if (is.na(s)) { : the condition has length > 1
```

Upon inspection, it seemed that the objects s2.GMOsubj.q, s2.scilit.q were missing. Because
the point of this section was to tabulate some quadratic analyses for the supplements and
my focus was on the results reported in the main body of the document, I opted to not try
to create these objects myself (potentially introducing other errors). I simply note the issue
here.

- I encountered a similar code error with the following:

```
> ## create the education controls table ##
> stargazer::stargazer(s1GMOEDU, s2EDU, s3EDU, s4EDU, type = "html", out =
"NHBALLstudiesEDU.htm")
Error: object 's2EDU' not found
```

- In the code section titled # QUADRATIC TESTS, STUDIES 1, 3, and 4 # the following code would not run:

```
> s4.GTsubj.q <- lm(s4$subjective.knowledge ~ s4$extremity.C +
I(s4$extremity.C^2))
Error in model.frame.default(formula = s4$subjective.knowledge ~
s4$extremity.C +  :
  invalid type (NULL) for variable 's4$extremity.C'
```

There was no s4$extremity.C variable created in the code. I've added a line to create this variable in Fraley Code Insert 11.

Time spent: 60 mins   minutes/hours

2. Computational Reproducibility of Reported Statistics `Errors found`

Is there a clear traceability of reported stats to code? Does the code output match what's reported in the paper? Are all reported statistics findable within the analysis code?

Errors or potential errors (maybe I made the error):

Extremity of opposition on self-assessed knowledge (Study 1)
* coef of 26 reported. Should be .26

Extremity of opposition on the z-scored knowledge difference score (Study 1)
* coef of 28 reported. Should be .28

Age Study 3
* Average age reported at 36.6 but is 39.6 in analyses (Study 3)

Other Notes:

Study 3. I didn't look in the supplements for potential analyses. But the broad conclusion reported in the paper is that the analyses in Study 3 were consistent with those reported for Study 1. (The major difference being the order in which questions were asked.) My run of the analytic code was compatible with that conclusion. I also redid the difference analysis with standardized extremity scores (for my own benefit) and the interpretation was consistent with the authors' conclusion.

Study 4. Exact results are not reported in the article, but the analyses are consistent with the general conclusions reached previously. For example, people who are more extreme tend to have lower scores on scientific literacy, higher scores on self-assessed knowledge, and a larger discrepancy between objective and self-assessed knowledge.

I did not attempt to recreate the figures exactly because I wanted to minimize the number of libraries I was using. I did create my own versions of Figure 1 and 2 just to eye-ball check things, and everything looked correct.

Time spent: 60 mis   minutes/hours

3. Data Processing Errors   No errors found
Are there substantive errors during the preparation or cleaning of data (e.g. duplication of rows during a merge) prior to substantive analyses and hypothesis tests?
Time spent: 60 mins   minutes/hours

I reviewed the data cleaning and re-coding elements for S1, S2, and S4 carefully. I did not identify any errors.

4. Model Misspecification   No errors found
Are there any consequential issues with the assumptions or the form of a statistical model (e.g., overfitting, wrong distribution assumption) used to describe data?
Time spent: 90 mins   minutes/hours

A minor concern (perhaps not an "error" per se): The intercept was not reported for the difference analyses; it was -1.28 based on the R analyses (Study 1). This may be worth reporting because the primary analysis concerns a difference between two scores and how that difference varies as a function of (non-standardized/centered) extremity. The coefficient of .28 implies that people who have more extreme scores also tend to have a larger difference between their self-assessed and objective knowledge. But, depending on where exactly the regression function crosses the zero-point on the outcome variable could matter for the interpretation of that result. A positive slope could indicate that more extreme people have less of a discrepancy if that intersection is shifted to the right and could indicate that more extreme people have more of a discrepancy if that intersection is shifted to the left. In other words, knowing where the crossover point is can be helpful for interpreting regression results when the DV is a difference score and the predictor is not centered/standardized.

In this case, the intercept is negative (-1.28) indicating that people with extremity values on the low end of the scale tend to have a negative discrepancy (i.e., underconfidence). So people with high scores could be approaching 0 or moving beyond that. One can add up the numbers, of course. But to get a better visual understanding of what was happening, I made some plots (Fraley Code Insert 4, 6, 9). The plots indicated that the person average in extremity had virtually no discrepancy between objective and self-assessed knowledge. People who were more extreme had a larger discrepancy (positive values indicating greater overconfidence). This is compatible with the authors' interpretation.

An additional check: Because the claim is that people who have more extreme opposition to genetically modified foods show a greater discrepancy between objective and self-assessed knowledge, I wanted to perform a simple-slopes test to see if, for example, that difference is significant at the extreme end (Fraley Code Insert 7). (Depending on the elevation of the regression line relative to the zero-point on the outcome, this may or may not pan out. But, given the additional plots I examined (see above), there was no reason to expect it not to.)

For Study 1, the difference is significant at both low extremes and high extremes. This latter finding is compatible with the general conclusion: That people with more extreme opposition to genetically modified foods show a larger discrepancy between their objective- and self-assessed knowledge (i.e., they report knowing more than they do).

I'm not sure what the former means, however: The predicted values imply that the difference goes in the *opposite* direction. People who do not harbor opposition to GMO foods report knowing *less* than they actually do know.

I don't think it is an "error" to leave this part out. But my intuition is that, to fully appreciate the implications of the findings, the account needs to be able to explain *both* ends of this continuum satisfactorily. Stated differently, the paper is written as if the "phenomenon" is that people with extreme opposition to GMO foods are misaligned in their knowledge (i.e., they think they know more than they do). But the take home point could have just as easily been that people who are not opposed to GMO foods are misaligned in their knowledge (i.e., they think they know less than they do). Obviously, both of these can be (and appear to be) true, given the data, so it seems incomplete to not address this later phenomenon. (Figure 2 seems to suggest that there is a greater alignment between objective and subjective knowledge among those who are low in extremity, but that is with a focus on the association rather than the discrepancy per se (i.e., the difference in standardized objective and subjective scores). Those could be positively correlated even if, on average, one kind of knowledge is higher (or lower) than the other.)

5. Erroneous/Impossible/Inconsistent Statistical Reporting No errors found

Are there inconsistencies between test statistics, degrees of freedom, and p-values? Are there implausible degrees of freedom between compared SEM models? Are there point estimates outside the confidence interval bounds?

Time spent: 10 mins   minutes/hours

6. Other Aspects Related to Data or Code Didn't check

Time spent: 0   minutes/hours

III. CLAIMS, PRESENTATION, AND INTERPRETATION

1. Interpretation Issues   Indeterminable

Throughout the entire paper, is there an incorrect substantive interpretation of data or statistical tests, causal inference issues, etc.?

- A minor concern: The authors commonly use within-persons, repeated-measures, or causal-like language to report between-person correlations. For example, in the Abstract, the authors write that "as extremity of opposition to and concern about genetically modified foods increases, objective knowledge about science and genetics decreases…" This "pull it and it gets longer" language implies that the authors are studying the way changes in one variable for a person result in changes in another for a person. But it would be more accurate to say instead that "People who were more extreme in their opposition to genetically modified foods were less likely to have objective knowledge about science and genetics." The latter is the "correct" language for between-person covariances. The language the authors used would be more appropriate if there were within-person covariances or experimental manipulations of knowledge or opposition.

- One of the core findings of this work (to oversimplify it) is that extremity of opposition to genetically modified foods is positively associated with self-assessed knowledge, but negatively associated with objectively assessed scientific knowledge.

To echo the Gelman and Stern concern raised in the Broad Summary: This conclusion is reached based on separate analyses (e.g., p. 252, 1st col, 1st paragraph) and noting that one coefficient is positive (and significantly different from zero) and another is negative (and significantly different from zero). An alternative way to examine this would be using mixed models where knowledge is treated as a within-persons variable (objective vs. self-assessed). This should not lead to different conclusions in this particular case; the associations are going in opposite directions, after all. But, that might not have been the case *a priori* and the analytic plan might not have accommodated this kind of issue well. (In the interest of being comprehensive, I did perform a mixed model test and, not surprisingly given the patterns already reported, there was a significant interaction between knowledge type (subjective vs. objective) and extremity in predicting knowledge scores (Fraley Code Insert 8).)

2. Overclaiming Generalisability   No errors found

Does the paper overclaim the generalisability of the findings with regards to stimuli, situations, populations, etc.? Is there hyping or overselling of the importance or relevance of findings?

3. Citation Accuracy Didn't check

Are there misrepresentations of substantive claims by cited sources? Inaccurate direct quotes? Incorrectly cited or interpreted estimates? Citations of retracted papers?

Time spent: 0   minutes/hours

4. Other Aspects Related to Interpretation   Indeterminable

Time spent: 0   minutes/hours

AUTHOR ЯESPONSE

Nick Light

Background

In April 2024, Phil Fernbach contacted me about an email he received from Dr. Malte Elson, who described the ERROR project and invited our participation in a review of our 2019 paper published in Nature Human Behaviour: Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019), "Extreme opponents of genetically modified foods know the least but think they know the most," *Nature Human Behaviour*, 3(3), 251-256. Phil asked if I would be interested in taking the lead on this process, as I was the coauthor team's keeper of materials, data, and code, with support from Sydney Scott.

Initially, I was somewhat reluctant to participate. I had visions of potential career-damaging errors being discovered, partly because I was a relatively junior PhD student at the University of Colorado when Phil and I started this project. Additionally, we had learned relatively late in the project's development that Sydney Scott, Yoel Inbar, and Paul Rozin were working on a very similar project with complementary results. The two separate coauthor teams ultimately combined, resulting in the five-person manuscript published in Nature Human Behaviour.

Without too much deliberation I decided to get over my reluctance. As someone interested in the history and philosophy of science, I believe in scientific principles including the revision of existing paradigms, confronting uncertainty, triangulating effects, dialogue, and replication to achieve a better understanding of the world. Given these beliefs, I couldn't decline participation. After all, this paper focuses on individuals' attitudes on a technology about which there is a scientific consensus (arrived at via decades of scientific knowledge production, critique, and reevaluation). Any rookie mistakes I might have made *should* come to light. I directed Dr. Elson to our materials in a public OSF repository and agreed to participate. Dr. R. Chris Fraley (University of Illinois Urbana-Champaign) was selected as the reviewer, and I received his comprehensive review in July 2024.

Noted Errors

I was pleased to learn that Dr. Fraley found no errors related to the paper's conceptual design, measurement, model specification, data processing, pre-registration consistency, generalizability claims, or statistical reporting. The most notable errors identified, in my opinion, were two typos in Study 1's reported coefficients, where decimal points were missing. Dr. Fraley also noted that some language describing correlational relationships comes dangerously close to implying causation. I take both issues seriously and commit to addressing them (and any other errors) in whatever manner the ERROR review team deems appropriate.

Overall Response

In my opinion, ERROR is a promising model for scientific quality assurance, and I appreciate the opportunity to participate. The idea of pairing rigorous post-publication reviews with incentives for both reviewers and authors is an interesting complement to traditional peer review processes.

I am especially hopeful about ERROR's treatment of error detection as an integral component of the research process that helps normalize critical examination of published work. This approach aligns with fundamental scientific principles of falsifiability and iterative refinement. Scientific findings warrant careful scrutiny, and ERROR demonstrates that thorough post-publication review can be implemented in a way that both identifies meaningful errors and maintains collegial scientific discourse. One of the things I love about academia is a culture that values both precision and constructive criticism (most of the time).

It was interesting to see which aspects of this particular paper were singled out for evaluation and scrutiny. There is a certain meta quality to this process, given the paper's focus on anti-science attitudes toward a technology about which there is a scientific consensus. In retrospect, I am proud of the paper we produced, and I hope readers update their priors about its findings (both "correct" and "erroneous") in light of Dr. Fraley's review and subsequent findings from other researchers. I would like to extend a special thanks to Dr. Fraley, Dr. Elson, and Dr. Cummins for their time, professionalism, and commitment to this pioneering endeavor.