ESTIMATING THE RELIABILITY & ROBUSTNESS OF RESEARCH

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

reviewed by

R Chris Fraley, University of Illinois

May 7, 2024

*For the sections below, indicate whether you discovered any errors using the dropdown menu. Describe the **errors** you discovered, the **methods** that you used to find them, and the **amount of time** you invested in the search. Refer to specific files to allow verification of your review. For the assessment below, make sure to check if authors have provided **supplementary analyses** as these may clear concerns arising from the (interpretation of) the primary analyses. If you have written code yourself for the review, please attach it to the report. Please indicate the version of software/packages you used to run the original code and/or your own code.*

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.?

Time spent: 90 mins minutes/hours

- 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?

Time spent: 5 mins minutes/hours

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