



ESTIMATING THE RELIABILITY & ROBUSTNESS OF RESEARCH

## AUTHOR RESPONSE

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of experimental social psychology*, 55, 110-125.

*response by*

Mina Cikara, Harvard University

*Review version 1.0 (2025)*

*Review template version 1.0*

*License: CC BY 4.0*

My co-authors and I are deeply grateful to Sørensen and to the ERROR team for this incredibly careful and constructive review of our data and code from our 2014 JESP paper.

The rigor with which this reproducibility review was conducted has yielded a tremendous resource to us, as authors, as well as the field—clean data along with functioning SAS code—which we can now share on OSF and link to the paper. We had yet to post the data and code publicly ahead of participating in this ERROR review, so this is quite a boon.

As Sørensen's report notes, once the updated code runs there are only a few minor discrepancies between the updated output and what appears in the published paper, none of which change any of the results or interpretations of the experiments:

- Minor discrepancies in Cohen's  $d$  across the studies: Sørensen says "The values acceptably close, and certainly close enough that the substantial interpretation does not differ. Note that both approaches here are approximations, and there is no ground truth to how to compute Cohen's  $d$ ."
- $F^2$  for experiments 3a and 3b: "The authors reported the effect sizes for this non-significant effect to be 0.0002 for both conditions, whereas I get -0.0002 for both" and "I get -0.0001 for both, and not 0.0001 as the authors report"
- Rounding differences 3b: "I get standard error 0.02493 for the first row (Empathy for negative events), which when rounded to two decimals becomes 0.02. It is reported to be 0.03" and "I again get 0.02493 which when rounded to two decimals becomes 0.02, but it is reported to be 0.03."
- Effect size in 4: An "effect size with value 0.0005 is reported for the howgood condition, but I get 0.0006 when running my revised SAS code."
- In Table 1, in the first row for Effect "Functional relations (FR)", the denominator degrees of freedom is given as 196. The correct value here should be 199.
- In the right paragraph on page 116, where marginal means are reported, there is some inconsistent reporting. For Schadenfreude out-group>in-group, it is stated " $t(94) = 3.06, p < .001$ ". However, the  $p$ -value here is 0.0029, which is not less than 0.001. A one-sided  $p$ -value would be  $0.0029 / 2$ , which is still not less than 0.001. Hence, correct would be " $t(94) = 3.06, p < .01$ ".
- Furthermore, there is an instance of inconsistent reporting in the same column on page 116. For empathy for positive events, out-group < unaffiliated, it is stated " $t(94) = -3.71, p < .01$ ". While this is mathematically correct, the  $p$ -value here is 0.0004, so it should have been reported as  $p < .001$ .

- On page 119, in the first paragraph in the section titled “Predicting how bad/good the participants felt” it is stated in parentheses “nor were any of the other lower-order interactions including the before-after feedback factor: all  $p$ s > .35”. While it is correct that none of the interactions were statistically significant, the term “bef\_aft \* resptype” has a  $p$ -value of 0.2609 in the output from PROC MIXED. Hence, the statement “all  $p$ s > .35” is not correct.
- In the last sentence of the first paragraph of section “Predicting how bad/good the participants felt” in the left column on page 120, a wrong number of degrees of freedom is reported for the  $F$ -statistic. It is reported as “ $F(1,67) = 5.96$ ”, but correct would be “ $F(1,70) = 5.96$ ”. In the second paragraph of the same section, a wrong number of degrees of freedom is reported for the difference in marginal means for positive events “ $t(170) = -7.00$ ”. The correct value should be “ $t(70) = -7.00$ ”.

We feel very fortunate to have been selected for this review and, again, are grateful to Sørensen and the ERROR team for this incredible service to our work and the field.