

Expression of Concern: Effort for Payment: A Tale of Two Markets

Psychological Science
 2021, Vol. 32(8) 1338–1339
 © The Author(s) 2021
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/09567976211035782
www.psychologicalscience.org/PS



This statement is an Expression of Concern regarding the article “Effort for Payment: A Tale of Two Markets” (Heyman & Ariely, 2004) published in *Psychological Science*. This Expression of Concern is prompted by some uncertainty regarding the values of statistical tests reported in the article and the analytic approach taken to the data. The corresponding author of the article and coauthor of this statement, Dan Ariely, attempted to locate the original data in an effort to resolve the ambiguities but was unsuccessful. Because the ambiguities cannot be resolved, we decided to issue an Expression of Concern about the confidence that can be held in the results reported in the article.

The ambiguity originally was brought to the attention of the Editor in Chief by Gilad Feldman, Hirota Imada, Wan Fei Chan, Yuk Ki Ng, Lee Hing Man, Mei Sze Wong, and Bo Ley Cheng. These researchers ran the article through *statcheck* (Epskamp & Nuijten, 2018), an R package that is designed to detect statistical errors (much like spell check and grammar check in Word). The program searches articles for statistical results, recalculates the values, and compares the reported and recalculated values to determine whether they match. *Statcheck* is now required for all articles published in *Psychological Science*. However, it was not available at the time of submission and acceptance of Heyman and Ariely’s article. The *statcheck* analysis conducted by Feldman and colleagues produced some discrepancies between the statistical values reported in the article and those determined by the recalculation. Gilad Feldman notified the editor of these discrepancies.

On receipt of Feldman and colleagues’ report, we reran the *statcheck* analysis in house. The report yielded 13 “errors” or mismatches. Five of the discrepancies were associated with Experiment 1 of the article, four with Experiment 2, and four with Experiment 3. The values reported in the article and those resulting from the *statcheck* recalculation are provided in Table 1. In seven of the instances of discrepancy between the reported and recalculated values, the *p* value of the test changed, but the status of the test—as statistically significant or not statistically significant—remained the same. That is, the test was reported as statistically significant and remained

so or was reported as nonsignificant and remained so. Thus, these discrepancies do not impact the interpretation or conclusions. In six of the instances of discrepancy, the outcome of the statistical test was different, resulting in what is known as a “decision error.” In these instances, there is a mismatch between results reported as significant and the results of the recalculations, which indicate that the tests are not statistically significant. These discrepancies, if the recalculated values are correct, would change the interpretation of the data in a manner that substantively alters the conclusions drawn from the research.

In addition to the issues raised by the *statcheck* report, there is a lack of specificity regarding the analytic approach adopted in the article. The authors did not provide a narrative description of the analytic approach, which featured seven conditions in Experiment 1 ($N = 614$ participants), five conditions in Experiment 2 ($N = 159$), and five conditions in Experiment 3 ($N = 90$). In all cases, the statistical tests were *F* tests; for all reported tests, the degrees of freedom are indicated as (1, 607), (1, 154), and (1, 84), for Experiments 1, 2, and 3, respectively. Although the original analyses cannot be located, the authors recall conducting the analyses using a planned contrasts approach in the *SuperANOVA* package, using all the data from all the conditions for the error term and the specific cells for the conditions. Note that this is a unique analysis approach, and the differences from the *statcheck* results could be due to the way this approach was carried out or reported.

Because the data for this article cannot be located, it cannot be determined whether the original reported values or those resulting from the *statcheck* recalculation are correct. Nor is it possible to more fully disambiguate the analytic approach. Given the ambiguities, the confidence we place in the conclusions drawn from the research is diminished. However, again, given the ambiguities, the Editor in Chief decided not to change the official publication record of the article through a Corrigendum. Instead, the corresponding author and editor are issuing this Expression of Concern and note that the differences between the values reported in the published article and the values recalculated through

Table 1. Comparison of Values Reported in the Article With Those Calculated by *statcheck*

Experiment	Statistics reported in article	Recalculated <i>p</i> value	Nature of discrepancy ^a
1	$F(1, 607) = 0.25, p = .81$ (candy condition, no difference between low and medium payment level)	.617	Significance-level difference; no change in pattern: n.s./n.s.
1	$F(1, 607) = 3.44, p < .001$ (interaction between compensation level and form of payment)	.064	Change in pattern: sign → n.s.
1	$F(1, 607) = 0.20, p = .84$ (low-level candy condition not different from control condition)	.655	Significance-level difference; no change in pattern: n.s./n.s.
1	$F(1, 607) = 3.36, p < .01$ (monetized-candy condition different from candy condition)	.067	Change in pattern: sign → n.s.
1	$F(1, 607) = 3.48, p < .001$ (monetized-candy condition, increase from low to medium payment level)	.063	Change in pattern: sign → n.s.
2	$F(1, 154) = 10.27, p < .001$ (cash condition, increase from low to medium payment level)	.002	Significance-level difference; no change in pattern: sign/sign
2	$F(1, 154) = 5.86, p < .001$ (interaction of compensation level and form of payment)	.02	Significance-level difference; no change in pattern: sign/sign
2	$F(1, 154) = 1.04, p = .30$ (low-payment condition, no difference between candy and control conditions)	.31	Significance-level rounding difference; no change in pattern: n.s./n.s.
2	$F(1, 154) = 12.53, p < .001$ (low-payment condition, difference between money and candy conditions)	.0005	Significance level rounds to same value; no change in pattern: sign/sign
3	$F(1, 84) = 2.41, p = .018$ (cash condition, increase from low to medium payment level)	.124	Change in pattern: sign → n.s.
3	$F(1, 84) = 2.52, p = .014$ (monetized-candy condition, increase from low to medium payment level)	.116	Change in pattern: sign → n.s.
3	$F(1, 84) = 3.11, p = .007$ (both cash and monetized-candy conditions, effort lower in the low-payment than control condition)	.081	Change in pattern: sign → n.s.
3	$F(1, 84) = 1.02, p = .31$ (nonsignificant difference in persistence between no-payment control and medium-payment cash conditions)	.32	Significance-level rounding difference; no change in pattern: n.s./n.s.

^aIn this column, “n.s.” indicates that the statistical test did not reach the conventional level of statistical significance; “sign” indicates that the test was statistically significant.

statcheck, and the lack of specificity regarding the analytic approach, undermine confidence in these data and the conclusions drawn from them.

Patricia J. Bauer
—Editor in Chief
Dan Ariely
—Corresponding Author

References

- Epskamp, S., & Nuijten, M. B. (2018). *statcheck: Extract statistics from articles and recompute p values* (Version 1.3.1) [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/statcheck/index.html>
- Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science, 15*(11), 787–793. <https://doi.org/10.1111/j.0956-7976.2004.00757.x>