Response to reviewer comments are written below the original comment in purple. We have removed sections reviewers wrote that did not include suggestions.

Reviewer: 1

Comments to the Author

First, the authors mention Bernoulli discussing outliers way back in 1777. I have no idea who Bernoulli is, so maybe he really did discuss the issue back then, but since he is then credited with coauthoring a 1961 report, it might behoove the authors to clarify exactly who he is and when he was writing (I am not familiar with Bernoulli myself.)

We appreciate the opportunity to ensure our work is as accessible as possible to broad audiences. As such, we have include a brief aside in line #XX indicating that Bernoulli was a prominent mathematician in the 1700s who is notable for his work in statistics and economics.

We agree it is odd to cite a work he is coined as authoring in 1961—but this is the English translation of a memoire from 1777 originally written in Latin. As none of the authors can read Latin, this seemed like the most appropriate reference to the work.

Second, the authors may wish to provide a concrete example of how outliers and their management can influence their results. Joe Hilgard just reported on how a video game experiment might have been massaged to produce statistically significant results by excluding "inconvenient" outliers. Here is his coverage of that issue: https://crystalprisonzone.blogspot.com/2019/06/comment-on-chang-bushman-2019-effects.html Providing this example can show how mismanagement of outliers really can have a deleterious effects on study results.

We agree that a concrete example would be beneficial in this context, and this is a great example. We have included a brief description with reference to Joe’s original post at line ##.

Related at one point the authors suggest outliers can reduce effect sizes, but they can increase them as well, particularly if the responses fall into a pattern. The authors might want to tap into some of the literature on mischievous responding.

We agree that discussing how outliers may also increate effect sizes is important and we have included this now at line ##. We think the work on mischievous responding is interesting, but do not think we could do it justice in this work without getting outside the scope of the current project.

In the results, the authors may want to reiterate what is meant by “statistical reasons” “participant reasons”.

We have included examples of these categorizations in the results section.

Reviewer: 2

Comments to the Author

On page 6, the authors mention that keeping outliers in the data can have negative consequences. For completeness, I would like the authors to add that also the unjust removal of outliers (extreme values that are just true random values from the underlying distribution) can have negative consequences when different groups are compared (experimental designs): a decrease in the variance and an increase in the Type I error rate (Bakker & Wicherts, 2014; this is another one than currently included in the paper). Nevertheless, this further illustrates that the detection and handling of outliers should be well described as is the main point of this paper.

We appreciate the suggestion and article and have included a sentence at line # to cover this important piece.

The authors choose to include year as a continuous variable. I wondered whether the results would are similar when year is included as a dichotomous variable.

This is an excellent question. We re-ran all analyses with year coded as a 2-level categorical variable in all analyses. The only analysis where the conclusion “changed” (in terms of a strict p<.05 cutoff) was regarding the mention of outliers for the industrial organizational journal. Given the continuous coding the result would be p=0.04, while with the dichotomous predictor the p=06. Therefore, we still consider our initial coding as representative for this dataset.