

Reply to decision letter reviews: #176-R&R

We would like to thank the editor and the reviewers for their feedback. We were pleased to learn that reviewers Dr. Toby Wise, Dr. Richard Brown, and Dr. Bjørn Sætrevik recommended an acceptance. Below we address the final minor point made by Dr. Katherine Fox-Glassman.

A track-changes comparison of the previous submission and the revised submission can be found on: <https://draftable.com/compare/gQFOaGXyajRP>

A track-changes manuscript is provided with the file: “PCIRR-RNR2-Fischhoff et al 1978 replication & extension main manuscript-track-changes.docx”

Summary of changes in response to feedback by Dr. Katherine Fox-Glassman

Abstract

- Changed the wording to an exploratory analysis regarding whether participants rated risks/benefits differently

Overview of the replication - Replication

- Added a sentence explaining that the t-tests are exploratory and do not test the original theory
- Added a sentence about the implications of finding differences

Discussion - Future Directions

- Added a brief paragraph to indicate there will be additional discussion around the results/implications of the t-tests and the exploratory direction of comparing the between-subject and the within-subject designs given the unified data collection.

Dr. Katherine Fox-Glassman wrote:

Thanks to the authors for their clear and thorough documentation of changes, both within the manuscript and in the Response to Reviewers.

On all but one of the points, the authors' explanations/edits/justifications do a good job of addressing my questions and suggestions in the first review.

The only remaining reservation I have is about the between-subjects t-tests. The authors write that they decided to run them "primarily due to the between-subjects design of the original study," and that they "will really only tell us if participants are rating risks differently than they are rating benefits. Unfortunately, this is all that can be done given the design."

I fully agree with that last sentence: that there's not really more to do in this area given the design, and that that's unfortunate. But I am wary of running analyses (any analyses, but especially this many) for the primary reason that it's the only analysis that *can* be done, in the absence of a theoretical reason to do so. The line "to make the most of the replicated design, we will also be conducting independent samples t-tests..." is one that would raise statistical red flags for me as a reader.

Is there literature that could guide us on whether to expect people to rate benefits differently than they do risks? (I mean in a group-comparison context; obviously there's good theoretical reason to look for correlations between how people rate risks and benefits.) Face-validity-wise, I don't understand what it would mean to say that "people rate the risks as greater than the benefits for technologies/activities A, B, and C, but vice versa for technologies/activities X, Y, and Z," unless the Ps had been instructed in how to quantify both risk and benefits on some common (or at least comparable) scale. In the absence of training, it seems possible that many Ps would instinctually judge risks based on human lives lost (or injuries, etc.), but judge benefits based on more economic criteria. Or that people could think about the risks as those to the people exposed, and the benefits as to society in general. Or that loss aversion could cause people to be more sensitive on the low end of the scale when rating risks, compared to how they use the same range of numbers when rating benefits. Or any number of other different uses of the same set of numbers when rating the two different constructs.

No matter the reason, if there are any differences in the assigning of values to "risk" and "benefit," that would render meaningless any direct comparison between the two constructs.

If the authors are set on including these t-tests in the study, they should include sound theoretical motivation for what kinds of differences they expect to see (effect sizes & directions), and for how those differences would be interpreted. But I would recommend omitting this whole analysis—the paper can easily stand on its own in terms of contributing to the field without the t-tests.