

Reply to PCIRR decision letter reviews #670: **Fox and Rottenstreich (2003) replication**

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item. We also provide a summary table of changes. Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

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

A track-changes manuscript is provided with the file:
PCIRR-S1-RNR-Fox-Rottenstreich-2003-replication-main-manuscript-trackchanges.docx
(<https://osf.io/v4htz>)

Summary of changes

Section	Actions taken in the current manuscript
General	<p>Ed/R1: Rationales were provided as to why the Big Ten Conference item remains unchanged for now in the reply. There was no change regarding this item in the manuscript. Some typos were corrected (Table 8 and section "Replication: As in the Original").</p> <p>R1: Rationales were provided as to why the Big Ten Conference item remains unchanged for now in the reply. There was no change regarding this item in the manuscript. We added a note on the deviation concerning the payment to Table 7.</p> <p>R2: Corrected typos and added requested references.</p> <p>Additional changes: We added a mixed ANOVA and test of order effects to the manuscript. We revamped our code, now with complete Rmarkdown and the use of ggstatsplot R package, with figures added to the main manuscript.</p>
Methods/Results	<p>We added a three-way ANOVA has been added to "Extension: A three-way mixed ANOVA for Studies 1a, 1b, and 3". We added a test of order effect d to "Extension: Order effect (Study1a Item3 and Study4)".</p>

Note. Ed = Editor, R1/R2 = Reviewer 1/2

[We note that we are not familiar with the titles and ranks of the reviewers, and looking for that information proves tricky. To try and err on the side of caution, we refer to all reviewers with the rank Dr./Prof. . We apologize for any possible misalignments and are happy to amend that in future correspondence.]

Reply to Editor: Dr./Prof. Romain Espinosa

Thank you very much for your submission. I have read your paper with great interest and received feedback from two reviewers. Given this feedback and my own reading of the paper, I recommend a revision to address the minor concerns that the two referees raised and that I also noted while reading your manuscript.

Please note that both Olivier and I commented on the Big Ten Conference item which is, in my view, the major element of the revision. Olivier's suggestion regarding the Bayesian statistics with the initial paper's results used as priors can be seen as a general comment about your overall replication project and for exploratory discussions. However, I understand that you want to stick as close as possible to the tests used in the original paper. Last, both Olivier and I commented on payments. While these comments do not challenge your design, they might call for a dedicated subsection in the paper or, at least, a bit more emphasis.

I put my comments below. (As always, consider them with caution and feel free to contradict them: I might be mistaken.)

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

—Recommender's comments—

.1. - In the original study, for Study 1, the authors have underlined some parts of the text. I checked in Qualtrics and, indeed, you have underlined these sentences. You might be willing to underline them in your paper as well (Table 5).

Thank you for the suggestion. We added the underline to Table 5.

.2. - You write that F&R did not mention that the Big Ten Conference had 11 teams and that they « assumed their participants would know that information ». In their paper, Note #3 discusses this point. The authors write that « most people believe the Big Ten has 10 teams, but, in fact, it has 11. ». It seems to me that the authors were aware of this uncertainty. They assumed that, overall, people would have ignorance beliefs 1/10 vs. 9/10 in the class formulation. They further add in the note: « Thus, for some participants, the class ignorance prior may have been 1/11 - 10/11. ».

Given this, you could just leave it as they did, couldn't you? Another possibility would be, at the end of the survey, to ask a question about it. Something like: in your opinion, how many teams participate in the Big Ten Conference? (You could tailor the ignorance prior at the participant's level.)

Response: Thank you for your feedback.

Our primary focus is to investigate whether participants' responses are influenced by the manipulation, and not whether they know a certain fact or not and how that knowledge impacts the effect. Given that regardless we need to make changes then in alignment with this core objective, we believe that incorporating an additional survey question to assess each participant's prior knowledge about the number of teams in the Big Ten Conference, and adjusting the ignorance prior accordingly for each participant, would complicate the study unnecessarily and shifts focus from the main point that this experiment was meant to test. Therefore, we felt that explicitly informing participants about the number of teams in the Big Ten Conference was the right approach and a more conservative accurate test addressing a weakness in the target's design that was communicated clearly and we can easily address with a very minor adjustment.

Moreover, Fox and Rottenstreich explicitly stated that the DJIA included 30 stocks in Study 1a (item 3) and Study 4. Consequently, we believe that explicitly mentioning that the Big Ten Conference comprises seven teams maintains consistency across our study items.

Action: Based on our primary objective, we kept the item unchanged at this time. We are ofcourse open to amending this given clear editorial guidelines.

.3. - I do not see any issue with the replacement of GM by IBM.

.4. - I might be too much of an economist here but, in my view, the lack of real incentives in Study 4 is an important deviation from the original study. (People are likely to exert more cognitive efforts if they are paid, especially for significant amounts of money. 10\$ here is a large amount given the time spent.) You mention this in Table 7 but not in Table 8. (By the way: you refer to Table 8 in Table 8. Is that a typo?)

Response: Thank you for the suggestion and catching the reference issue.

Concerning the absence of real incentives in Study 4, we have explicitly highlighted this as a departure from the original study and explained our reasons in Table 7 of the Method section. Additionally, we note that Fox and Rottenstreich did not provide any details regarding what they did and the proportion of participants who received actual rewards. They wrote:

The participants in Study 4 were Duke M.B.A. students who were told that some respondents would be selected at random to have their decision honored for real money.

Therefore, while some won some money, some did not, we do not know the exact details. We believe our compensation mechanism is more fair. Prolific participants tend to be very competent in tasks similar to those of taking part in a study, and they regularly do these tasks for money. Our compensation is clearer, there is no luck involved, all are compensated. If indeed money is associated in some way with the exertion of cognition effort, then - if anything - our participants on Prolific with a clear certain compensation and reputation based system might be more likely to exert that effort.

Action: The typo in Table 8 has been corrected to refer to Table 7.

- « while others were rounded to one decimal place —> 0.3 ». If I am not mistaken, they use 0.03 not 0.3.

Response: Thank you for your careful review.

Action: The typo has been corrected. The text “0.3 for item 3 of Study 1a” has been updated to “0.03 for item 3 of Study 1a” under the section titled “Replication: As in the Original” - Data Analysis Strategy - Method.

- I think that there are some rationales for the choices of the authors of the original paper for the classification. For item 1 of S1: the chances are 1/7, which is 14.29%. So, assuming participants had to report percent integers, they could choose 14 or 15. So the authors considered both answers as correct. For the sports items: the probability is 1/10 (=10%) if they believe that there are 10 teams and 1/11 (=9.09%) if they think that there are eleven teams. So, they considered both answers as correct. For item 3, I agree that they should have considered 3% and 4% if they used the same method as for item 1. For Study 3, they can have used the same rule as for item 2 from S1 (in the case people thought there were 10 teams).

—> In my opinion, the original approach is very strict in determining which answers fit the class partition beliefs. People are not very good when it comes to probabilities. I think that your approach is more appropriate in this regard. Personally, I would have preferred a more continuous measure of the closeness/distance to one theory relative to the other. But this would bring you too far from the replication. (Maybe as exploratory discussions?)

We appreciate your insights and the support for our chosen methodology, as well as your suggestion for more robust tests.

We indeed are analyzing the estimation not just as whether they fit with the ignorance priors or not, but also as continuous measures. This partly addresses the flexibility of what ignorance prior threshold or range to set. Additionally, we have a planned exploratory three-way mixed ANOVA detailed in section “Extension: A three-way mixed ANOVA for Studies 1a, 1b, and 3”, which would be the more robust analysis of the continuous measures.

Reply to Reviewer #1: Dr./Prof. Olivier L'Haridon

The replication study is carefully designed and planned; I have only minor comments.

Thank you for the positive and supportive opening note and the constructive feedback.

.1. My first comment refers to Study 1a, Item 2. Due to the changes in the league, the ignorance prior is now 1/7, as explained by the authors. My point here is that the league was chosen because it included a genuine ignorance prior of 1/10. An alternative for the replication would be to use another league with a genuine prior of 1/10 instead of replicating the design with the initial league but changing the prior.

Thank you for your suggestion. Our primary focus is to investigate whether participants' responses will be influenced by partition priming. In line with this objective, we believe that the specific number of ignorance priors is not critical. For our replication, we aim to closely adhere to the items outlined in the target article rather than introducing new ones with a different league. Furthermore, we are analyzing both ignorance priors and the continuous measures, see our reply to the editor above..

.2. My second comment refers to the payments. One dollar in 2002 is equivalent to approximately 1.65 dollars in 2024. From the replication design, I understand that payments will be anchored on the minimum wage per hour. The authors should provide a better justification for not simply replicating the initial payments, adjusted for inflation (or some other purchasing power parity index).

Response: Thank you for your comment. Our participant recruitment will be conducted through Prolific, which mandates a minimal hourly reward requirement. Fox and Rottenstreich employed varying compensation schemes across the four studies we chose to replicate, as detailed in Table 4 of the Method section in the manuscript. Notably, the original studies exclusively targeted university students, affording Fox and Rottenstreich greater flexibility in compensation adjustment. In contrast, our participants are from the general US population, and we adhere to and aim to exceed the required minimum federal wage and for the payment rates set by Prolific.

Action: In Table 7 of the Method section, we added compensation as a deviation from the procedures outlined in the target article and included the justification.

.3. My last comment is a simple suggestion on statistical methods. An additional statistical method, if relevant in the current context, would be to use Bayesian statistics (e.g., Bayes factors) with the initial paper used as a prior.

Thank you. In our revised code and manuscript we included the use of plots based on the R package ggstatsplot which reports the Bayes Factor.

We believe that using the initial paper as a prior is not ideal and would introduce bias to the outcome. That reliance on a subjective factor giving the one original finding weight is something we consider to be both a strength of Bayes Factor but at the same time a major weakness. Depending on the outcome and one's own view on the phenomenon and the state of psychological science, some may argue that we introduced unneeded bias into what should ideally be an objective test of whether original findings hold without the original findings impacting that evaluation.

From our point of view, we would rather try and be as neutral as possible, and so we opted to use the default Cauchy of 0.707 meant to serve as a neutral assessment of the likelihood of finding the effect.

Reply to Reviewer #2: Prof. Don Moore

I like this project and think the authors have developed a worthwhile plan.

Thank you for the positive and supportive opening note and the constructive feedback.

For my PCI RR review, I have been asked to assess 5 things:

1. The scientific validity of the research question(s).

Here, I find the effort worthwhile. The original result is interesting, important, and influential.

2. The logic, rationale, and plausibility of the proposed hypotheses

Yes, this obviously derives directly from Fox and Rottenstreich's (2003) paper.

3. The soundness and feasibility of the methodology and analysis pipeline

Yes, I find the plans feasible and methodology sound.

4. Whether the clarity and degree of methodological detail is sufficient

No plan is perfect, but this one is pretty darn good.

5. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

Yes, I think the plans are sufficient. However, I will confess that I did not download and go over their survey materials with a fine-toothed comb (or any comb for that matter).

Thank you for your feedback. We appreciate the effort you have dedicated to reviewing our manuscript.

.6. A few other, smaller notes:

- **In the PCIRR Stage 1 snapshot: LeBel's name is incorrectly capitalized: "Lebel"**
- **Zwaan et al. (2018) is cited but does not appear in the references.**
- **Drop "the" from "the Fox and Rottenstreich (2003)'s studies."**

Response: Thank you for your detailed review. The typos have been corrected as noted.

Action:

1. We corrected the spelling of LeBel's name throughout the manuscript.
2. Zwaan et al. (2018) has been added to the reference list.
3. We revised the phrasing from "the Fox and Rottenstreich (2003)'s studies" to "Fox and Rottenstreich (2003)'s studies" in the final paragraph of the section titled "Choice of Study for Replication: Fox and Rottenstreich (2003)."