

## **Reply to decision letter reviews: #187 revise and resubmit 2**

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold while our answers are underneath in normal script.

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

**A track-changes manuscript is provided with the file:  
"PCIRR-RNR2-Soman 2001-Replication-Manuscript-v4-G-trackchanges.docx"**

### Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

<b>Section</b>	<b>Actions taken in the current manuscript</b>
General	N/A
Introduction	R2: expanded the introduction (3rd paragraph)
Methods	R1: added a table summarising methodology differences (Table 2) R2: doubled the sample (1212 recruitment with a target of 1030 final N) in order to be able to run order effects analyses. Updated power analysis section. R2, R3: Implemented additional familiarity questions and comprehension check questions, which we use as exclusion criteria
Results	R2: committed to order effects analysis and included a table (Table 7) summarising our findings
Discussion	N/A
Reporting	N/A
Supplementary materials	R2, R3: added the familiarity questions and comprehension check questions

*Note.* Ed = Editor, R1/R2/R3 = Reviewer 1/2/3

## **Response to Editor: Prof. Chris Chambers**

**Three of the five reviewers kindly returned to evaluate your revised manuscript. As you will see, most of the major issues have been addressed and we are moving closer to Stage 1 IPA. However, there are some remaining points to address concerning the literature review, exclusion criteria, and planned statistical analyses, and the reviewers offer a range of constructive suggestions for resolving these issues. I look forward to seeing your response in a further (hopefully final) Stage 1 revision.**

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

## **Response to Reviewer #1: Prof. Dilip Soman**

**My review of the revised manuscript is short because the authors have addressed the issues I had raised in the last round.**

**The only other suggestion I have is merely a suggestion (but one that I've found to greatly increase the readability of replication papers) is to include an additional table that compares the procedures and analysis plans across the studies (basically, the equivalent of Table 2, but for procedures and analysis). Rows might include: a) Design, b) Method of recruitment, c) Whether study was done independently or as part of a package, d) Mode of data collection (online, paper-pencil), e) Implementation of Manipulations etc. All of this information is currently in the manuscript and it might be good to pull it all together into a table. Also, this might be a good template for others in the field to follow!**

**All the best, and I look forward to seeing the project move further along.**

Thank you for the overall positive feedback and constructive comments in both iterations of the feedback. We added a new table, namely Table 2 (note that table numbers have been changed now), that covers the key dimensions across which the studies differ to allow the reader to get the methodological context at a glance.

## **Response to Reviewer #2: Dr./Prof. Christopher Olivola**

**The authors have only partially addressed some of my original comments, and failed to adequately address my core comments/concerns. Below, I go through my original points, and comment on the authors' responses to each one**

**Comment 1) “[...] manuscript is poorly written [...] I sincerely hope the authors will make an effort to proof-read their word before they submit it for review.”**

**=> The authors say that they have addressed grammar/spelling errors. I certainly hope so, but I didn't go through the trouble of proof-reading the manuscript, so I'll simply have to trust that the revised version is largely free of errors.**

**Comment 2) “The authors should cite and discuss other papers (besides Soman, 2001) that have previously tested (and found) sunk-cost effects for time (e.g., Bornstein & Chapman, 1995; Bornstein, Emler, & Chapman, 1999; Frisch, 1993; Navarro & Fantino, 2009; Olivola, 2018; Strough et al., 2008). This is important, since these other papers also speak to whether (and to what extent) there is a sunk-cost for time. In fact, I would suggest the authors provide a table that summarizes these other papers and, for each one, what they found (e.g., whether they found a sunk-cost effect of time).”**

**=> The authors have largely ignored and seemingly chosen to “dodge” this comment.**

We did try and address the suggestions in our previous revision, and we understand that we could have done a better job explaining what we did. We are glad to include citations that we feel are relevant and make the manuscript stronger.

We previously went through the citations you provided in detail and considered each of them. Three of the citations have already been included, and two of them did not have any scenarios involving sunk time costs and so we did not think those were relevant to include.

We added the one suggested citation that we seemed to have missed, the Bornstein and Chapman (1995), now included in paragraph 3 of the introduction.

**...This is extremely puzzling to me, since this comment is (i) easy to address, and (ii) important for reasons I previously tried to explain, but will now elaborate on further: The authors argue that “the intended scope for this replication is rather narrow”, but this argument fails on several grounds. First, replicating an effect does not absolve the authors of their responsibility to properly acknowledge and discuss the relevant literature. In this case, the relevant literature clearly concerns a (possible) sunk-cost effect for time (vs. money). Therefore, the authors should cite (and summarize the findings of) any/all prior papers that have tested such an effect, even if it was not the only/main goal of a given paper. Such a discussion is important, as it gives readers a proper sense of what the existing evidence is concerning time sunk-cost effects (or lack thereof), as well as possible boundary conditions, etc.**

**Second, I’m not sure how “narrow” the authors want the scope to be, but I am sure that the goal of a replication (and science, more generally) is not merely to replicate a particular study, but rather to examine the (in)existence of an effect (or set of effects) more generally. Therefore, it is not the Soman (2001) studies and findings per se that are interesting and important, but rather the establishment of a sunk-cost effect for time (or failure to do so). Consequently, the authors cannot argue that discussing other prior studies that tested and found or failed to find time sunk-cost effects is beyond the scope of their manuscript.**

We appreciate this suggestion, and we see the importance of a summary of the literature. We followed what we saw as common in similar other replication manuscripts we and others conducted on PCI-RR and social psychology journals (which we provided examples for), which focuses the scope of a replication on the target article for replication. We understand that we may differ in our opinion on what our role and responsibility should be in a replication project. We see our main responsibility to be about conducting a rigorous empirical close replication of one target article.

We think it important to include several of the key citations in the follow-up literature, and we included several of those in our introduction, yet the scope of a replication as we defined it for our investigation is and has been encouraged by PCIRR is far more narrow than a comprehensive review of the literature covering any/all papers published since and analyzing their effect sizes. Furthermore, the task of summarizing a literature is anything but easy, and should have a dedicated systematic review/meta-analytic effort to address it comprehensively and make sure it lives up to its importance. Doing that in a replication manuscript runs the risk of distracting readers from what the manuscript was meant for - an empirical replication of one classic article in that literature.

We think it is important to align expectations here. We understand that you have a much wider aim of verifying and summarizing the entire literature beyond Soman (2001), yet we cannot address that aim in this effort. Our aim is much humbler and very specific in regards to focusing on the Soman (2001) article. Regardless of our findings in our replication of Soman (2001), we make no claims regarding the broader literature. We feel the focus on this article is interesting and important enough.

We also addressed the issue about the importance of replications of a single target in our previous reply:

At this point in time we believe there is already a general consensus regarding the need for replications, and “making replications mainstream”, without the specific need to justify a replication of a well-cited impactful paper that has not been subjected to independent pre-registered/Registered Report direct replications.

To address this point we now also include an explicit section about this need, with some citations of the work on that need:

We aimed to revisit the classic phenomenon and examine the reproducibility and replicability of the classic findings by replicating the studies and improving the design with extensions. Following the recent growing recognition of reproducibility and replicability in psychological science (Brandt et al., 2014; Open Science Collaboration, 2015; Nosek et al., 2022; Zwaan et al., 2018), we embarked on a well-powered pre-registered replication and extensions of Soman (2001).

Our introduction in this manuscript, our focus on a single target rather than an entire literature, and our justifications for the importance of replications is similar to other PCIRR manuscripts from our team that received in-principle acceptance. Examples:

1. Yiu, S., & Feldman, G. Revisiting the psychological sources of ambiguity avoidance: Replication and extensions of Curley, Yates, and Abrams (1986).  
[\[In-principle Acceptance\]](#) [\[Preprint\]](#) [\[Open materials/data/code\]](#)
2. Lee, S., & Feldman, G. Revisiting the link between true-self and morality: Replication and extensions of Newman, Bloom and Knobe (2014) Studies 1 and 2.  
[\[In-principle Acceptance\]](#) [\[Preprint\]](#) [\[Open materials/data/code\]](#)
3. Lu, S. & Feldman, G. Associations of fear, anger, happiness, and hope with risk judgments: Revisiting appraisal-tendency framework with a replication and extensions of Lerner and Keltner (2001).  
[\[In-principle Acceptance\]](#) [\[Preprint\]](#) [\[Open materials/data/code\]](#)

4. Li, M., & Feldman, G. Revisiting diversification bias and partition dependence: Replication and extensions of Fox, Ratner, and Lieb (2005) Studies 1, 2, and 5. [[In-principle Acceptance](#)] [[Preprint](#)] [[Open materials/data/code](#)]
5. Yeung, K., & Feldman, G. Revisiting stigma attributions and reactions to stigma: Replication and extensions of Weiner et al. (1988). [[In-principle Acceptance](#)] [[Preprint](#)] [[Open materials/data/code](#)]
6. Zhu, M. & Feldman, G. Revisiting the links between numeracy and decision making: Replication of Peters et al. (2006) with an extension examining confidence. [[In-principle Acceptance](#)] [[Preprint](#)] [[Open materials/data/code](#)]
7. Li, M. & Feldman, G. Revisiting mental accounting classic paradigms: Replication of the experiments reviewed in Thaler (1999). [[In-principle Acceptance](#)] [[Preprint](#)] [[Open materials/data/code](#)]

**Third, were the authors serious (but mistaken, in my opinion) in trying to argue that their only goal should be a very narrow/specific attempt at a replication of Soman’s studies, then their proposal would have to be considered a failure, since they are using a different design, different subject population, etc. In other words, the authors can’t have it both ways: they cannot claim to be narrowly focused on specifically replicating Soman’s originally studies (as a supposed license to ignore other prior papers that tested time sunk-cost effects), while at the same time carrying out a replication that alters (i.e., departs from) many features of Soman’s original studies, and thus would generalize the testing of those findings beyond the paper.**

We see these points as separate and we see no contradiction in our aims. What we are attempting is a very common replication effort. Replications will deviate from the original, in time, sample, context, and sometimes in design. That is not a failure or a weakness, that is a strength. We are using best practices and most up to date categorizations regarding replications (e.g., LeBel et al., 2018 and LeBel et al., 2019) to transparently document deviations and adjustments and reflect on those with humility and caution. As you’ve seen yourself, in this case we are fortunate to have the author of the target article agree with these aims and claims.

Using these criteria, we categorized and justified our categorization as “close replication” and explained that in detail (see “Replication classification” supplementary as well as main text).

**In sum, the authors can either (i) carry out a very narrow (i.e., exact) replication of Soman’s studies, which means they would need to replicate the original design features and collect data from a similar population OR (ii) they can generalize their replication to be about a time sunk-cost effect more**

**generally, which allows them to depart somewhat from the original study but also requires that they acknowledge (and not ignore) other prior studies (including the ones I cited, above, as well as those provided by Review #4) that tested time sunk-cost effects. In my opinion, the latter (ii) is far more interesting and will make a greater contribution. Regardless of what they decide to do, the authors cannot have it both ways. Just to be clear, I was not suggesting that the manuscript should shift to becoming a thorough and detailed review of time sunk-cost effects. However, I think it is well within the scope of the paper to at least offer one table (and maybe even a graph plotting effect sizes) summarizing the results of prior studies involving time sunk-cost effects, and at least one paragraph summarizing what those prior studies found (overall).**

Our aim is the first of these two, and what we set out to do best we can using best practices for replications. Our replication includes and strengthens the original's methods, and we document all adaptations and will interpret all findings in accordance. All of our adaptations are typical to this literature and for replications, and we conducted those and published after extensive peer-review with similar adjustments and properties dozens of times before. We believe that they offer an important contribution, both revisiting the original findings with a reproduction and replication, and with several additional extensions that go beyond the original.

We hope that this will inspire more interest in this domain and would be happy to see others follow up on the directions you outlined.



**Comment 3) “Having the same participants complete all 3 studies in a single session is problematic, as it may cause spillover effects, amplify demand effects, etc. The authors should consider randomly assigning participants to one of the 3 studies (not all 3). Or, at the very least, the main analyses should only focus on the first study that each participant is assigned to (and subsequent analyses can look at all 3 studies within-participant).”**

**=> It looks like--though this is not entirely clear from their response--that the authors agree to carry out analyses on only the first set of studies/conditions that participants see. This should be done regardless of what the within-subject analyses find. In fact, as I previously indicated, the between-subjects analyses (on the first scenario subjects see) should really be the \*main\* analyses that they carry out (i.e., the within-subject analyses should be the additional/bonus ones; not the other way around).**

We have now committed to re-running all analyses in the paper split by whether Study 1 or Study 2 were presented first - see “Order effects between studies” section as well as Table 7. We have doubled our target sample size from 600 to 1212 ( $515 \times 2 + 15\%$  planned exclusion rate).

We wish to note that the request for running analyses only on the first displayed study is a far more conservative and strict test of the original’s findings, because of the implications on power, reducing the likelihood of detecting the effect, rather than increasing it. Our experience has been that order had very little to no implications in our many other PCIRRs and replications. In the 7 in-principle accepted PCIRR we conducted and listed above using a similar unified design, we found support for most of the findings in all 7 replications. Order can be easily tested for as moderator, and if there are no order effects, there is little reason to discard doubling the sample.

**Also, it is critical that participants not know, in advance, that they will be completing multiple studies/conditions on the same question/domain. That is, while participants may be told that they will be completing various studies, they should approach the first sunk-cost scenario without expecting to see additional sunk-cost scenarios (i.e., these should come as a surprise). Otherwise, there could still be spillover effects on the first scenario if subjects know they will be asked about similar things in subsequent parts of the survey. Also, every sunk-cost scenario should appear on a separate (web) page, so that participants’ response to the first scenario is completely unaffected by the next one they are presented with.**

We shared [a preview to our Qualtrics in the manuscript](#), and the Qualtrics export files are also made available in the shared Qualtrics in the OSF folder.

To make it clearer, participants are given an outline of the studies at the start which:

“This survey involves 3 decision-making tasks, the first 2 presented in a random order. Please read the scenarios and answer questions carefully.

In each decision-making task, you will be presented with a short hypothetical scenario.

There are no “right” or “wrong” opinion answers, so please state your opinion as honestly as possible.

After that, there you will be asked some questions about your participant and provide demographic information. ”

Each study is implemented in separate Qualtrics blocks, which appear in separate pages.

**Finally, the authors need to make sure that they increase their sample sizes so that they have sufficient power (even) for the between-subjects analyses that only focus on the first sunk-cost scenario that each subject sees.**

We were already well powered to detect weak effects, far weaker than those reported in the original. Aiming to alleviate any further concerns, we went ahead and doubled our planned sample again.

**Comment 4) “Another concern, which may lead to a failure to replicate the effect, is that experienced MTurk participants may have been exposed (and some repeatedly) to sunk-cost studies, and this may hinder the effect. The authors should therefore consider limiting the study to MTurk participants who have had relatively little experience (e.g., fewer than 100 MTurk studies completed).”**

**=> The exclusion question/item that the authors propose to use seems far too vague (“Have you ever seen the materials used in this study or similar before?”), especially since it appears at the very end of the survey. The authors should (also) ask more specific recognition questions, along the lines of “Have you participated in any other study/studies that presented a scenario involving a decision to select a project that someone had already invested time or money in? Have you participated in any other study/studies that presented a scenario involving a decision to select an item that someone had already spent time or money to obtain?”**

MTurk participants who have done fewer than 100 studies are too new to the system, and lack experience not only with sunk cost, but in general in being MTurk workers and taking part in studies. Also, they still lack the commitment to high quality work on the platform (with little implications for their reputations), which raises a host of other quality issues. The issue of familiarity can be address by other means.

We added an additional “familiarity check” after Studies 1 and 2, in which we ask participants if they have encountered the scenario before. Those who answer that they have will be excluded.

Details can be found in the supplementary materials as well as on the Qualtrics preview link and Qualtrics exported files.

**Comment 5) “I don’t understand the distinction that the authors are trying to draw, here. Chi-Square tests also evaluate whether likelihoods vary across conditions, so the authors are mistaken if they suggest otherwise. I suspect they meant something else, but that it did not come across clearly in their writing.”**

**=> My confusion stemmed from the fact that the authors seemed to (unintentionally) imply that Chi-Square and Logistic Regression modeled different outcome variables. To avoid confusion, the authors can add (to the revised sentence they propose) that they propose testing an interaction, and that this can only be done with Logistic Regression (or, at least, not with a standard Chi-Square test).**

We revised the manuscript to make this more explicit:

“To address this question, we conducted logistic regression analyses for Study 2 for both the original and the replication data as it allowed us to test the interaction effect.”

## **Response to Reviewer #3: Dr./Prof. David Ronayne**

**I thank the authors for their responses.**

**Below I detail some residual issues and suggest remedies. Other than those, I am happy for the authors to proceed with their study.**

Thank you for your valuable feedback in both iterations of revisions.

**Outstanding issue 1: exclusion criteria. Your response to my and Johanna Peetz's comments clarify things greatly (which was needed - and overviews should be in the final paper even if summarized/footnoted). I now see the practices of the services you use (CloudResearch/Turkprime) help improve data quality, which is of course very welcome (as is your explicit description thereof).**

**However, I fail to see how that means you should not include explicit and credible criteria within your experiment (i.e., those that allow for how inattention/poor quality may interact with your specific study). The only thing I see that responds to this concern is that you highlight you ask subjects if they will pay attention (yes/no response options), and then expost whether they were serious (Likert scale) and understood what they read (Likert scale). Asking someone if they intend to pay attention or understood is not the same as checking they did pay attention or did understand, and is not convincing at all.**

**It would be convincing to include simple and explicit criteria based on hard data such as: a minimum time spent on the experiment's crucial pages (e.g., excluding those who spend less than X seconds on a page); or explicit attention checks (questions subjects know the answer to only if they paid attention - these should be standalone questions aiming to assess attention only, not those that are also part of the experiment, you could also mention at the start that they should pay attention because you will ask comprehension questions).**

**If your data are guaranteed to be of high quality as you argue strongly in your responses, then they will not bite: your subjects will all pass such criteria and none will be excluded. But your work would be stronger and your results more trustworthy.**

Thank you for the detailed suggestions. We have already implemented some comprehension checks for the education manipulation, which were not present in the original, and already informed participants that comprehension checks will be conducted and they should be paying attention.

Although we are reluctant to include any hard criterion regarding time spent on a page, given its arbitrary nature (e.g. we consider a participant to be legitimate even if they have went to the bathroom in between studies or during the reading of a scenario), we agree that we can include further comprehension checks to test whether participants understood the critical information in the scenarios. We have implemented these checks after each of the studies in Studies 1 and 2 and before asking them if they are familiar with the scenario (see our response to Reviewer 2) as we do not want to detract their attention from the main response. If we were to insert these checks between the scenario and the eventual response, some would consider that to be a major deviation from the original. Instead, we opt to exclude participants post-hoc if they have not understood the scenario. Although that could potentially strain our resources as some participants might be rejected, we are already overpowered (given our new target of 1212 PPs, see our response to Reviewer 2) so that should not be an issue.

The content of the comprehension checks can be found in the “Comprehension checks subsection” under “Materials used” in the supplementary or on the preview link here: [https://hku.au1.qualtrics.com/jfe/preview/previewId/af9ed4a0-f30e-4088-8408-07e4b05a8ecd/SV\\_bNSYF5TiOnh8VrU?Q\\_CHL=preview&Q\\_SurveyVersionID=current](https://hku.au1.qualtrics.com/jfe/preview/previewId/af9ed4a0-f30e-4088-8408-07e4b05a8ecd/SV_bNSYF5TiOnh8VrU?Q_CHL=preview&Q_SurveyVersionID=current) .

Participants who answer any comprehension check incorrectly were excluded.

**Outstanding issue 2: Please state explicitly (instead of requiring the reader to go and open your Qualtrics project file): can subjects go back to previous pages of the experiment? In your case it would seem important that they cannot go back (including for explicit attention checks).**

We do not have a previous arrow. We made that explicit in the manuscript:

“Throughout the study, participants could not go back to previous pages.”