Reply to PCIRR S2 decision letter reviews #775: Baron and Szymanska (2011) replication and extension

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/AlKEzuxCPDgX (https://osf.io/pn5va)

A track-changes manuscript is provided with the file: "PCIRR-S2-RNR-Baron-Szymanska-2011-replication-extension-main-manuscript-trackchanges.docx" (https://osf.io/wusdv)

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:

Section	Actions taken in the current manuscript
Abstract	R1 (Minor 2): Directions of results made explicit
Introduction	R1 (Minor 4): Definition of "efficiency" made explicit R1 (Minor 6): Table 1 N/As changed to hyphens R2 (Minor): Caviola cite description updated
Methods	R1 (Main 2): Post-hoc power analysis added R1 (Main 3): Mentioning that scales were not randomized
Results	R1 (Main 4): Replication and Extensions headings updated R1 (Minor 9): One sample t-test midpoint change made explicit
Discussion	R1 (Main 1a): Plain language result interpretations added R1 (Main 1b): Discussion about difference in effect sizes for Hypothesis 1 R2 (Main 3): Hypothesis 6 external validity discussions added
Conclusion	R1 (Minor 2): Directions of results made explicit

Section	Actions taken in the current manuscript
Overall changes	R1 (Main 1c): Effective -> efficient when describing Hypothesis 3 R1 (Minor 3): Hypothesis 2 description updated R1 (Minor 5): Hypothesis 7 renamed to “External funding”
Supplementary materials	R1 (Main 3): Update to the reason why scales were not randomized

Note. 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 submitting your Stage-2 manuscript to PCI-RR.**
> **I am sorry for the long delay since your resubmission.**
> **The two reviewers who evaluated your Stage-1 manuscript kindly agreed to read your completed manuscript and gave precious feedback. You can access their reports below.**
> **As a recommender, I believe that you closely followed your Stage-1 manuscript and congratulate you for this useful replication work.**

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

> **I think that the miscalculation of the statistical power in the Stage-1 is slightly disappointing, but mistakes happen very often and you clearly reported the issue. I appreciate the transparency. On this point, I concur with Amanda's opinion and would also suggest that you report the corrected statistical power. Please note that the corrected power analysis should not be conducted using the effect size estimated from the data that you collected. As far as I am concerned, I would be satisfied with you using the same method/script as in the Stage-1, correcting for this mistake, and reporting what power this would have yielded. This technical mistake is not a concern for the recommendation of the manuscript, of course.**

Thank you for the feedback. We conducted and now report a sensitivity analysis and adjusted the note at the end of the power analysis section to the following:

> After data collection in Stage 2, we noticed an oversight. We initially conducted our power analysis based on an alpha of .05, and based on the Stage 1 peer review recommendation we adjusted our alpha to .005, yet did not update our power analysis. This did not have much impact, as we targeted d = 0.2 which was an extreme under-estimation. A sensitivity analyses of a sample size of 350 per study (1403/4), power of 0.95, and an alpha of .005, one-sided, showed we were powered to detect a one-sample effect of d = 0.23, and a paired-samples effect of dz = 0.23.

> **Regarding the other comments of the reviewers, I think that some of them could really help you improve your manuscript and make it more reader-friendly and/or more informative for the readership.**
>
> **For instance, I share Amanda's view that the paper would benefit from more intuitive presentations of the findings (in plain language, not only in statistical language). Another example is the use of the word « Efficiency », which might be associated with different understandings. In my view, you explain what you understand with this word on Page 8.**

Thank you for the feedback. We provide a summary of the findings in plain language in the discussion, yet we understand the need to also revise the presentation in the results section.

In our revision, we worked on making the discussion more explicit and in clearer language that connects between the tests and the hypotheses. Please see all of our revisions to the results section.

> **Last, there are several comments that address the generalizability of your findings and/or that seek to make sense of them. I think that some of them could be interesting to discuss (the reviewers' comments show the type of questions your readership will ask, so you might want to anticipate it and extend your discussion section). I leave it up to you to decide to which extent to enrich your manuscript in this regard. Most importantly, I will not condition the recommendation on running additional analyses as you did what you committed to doing.**
>
> **I am looking forward to your revised manuscript and congratulations again for your work.**
>
> **Thank you very much for considering PCI-RR for your work.**

Apart from Dr./Prof. Berman's initial suggestion on including additional analyses, we by and large have included the suggested points of discussion in this revision.

Reply to Reviewer #1: Dr./Prof. Amanda Geiser

> **Thanks for giving me the opportunity to review this submission again at Stage 2. Overall, this version of the manuscript looks good to me, and it's great to see that most of the replications were successful.**

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

> **I will list my main comments first, which are based on the factors that PCIRR recommends reviewers to focus on at Stage 2. Then below I will list some more minor comments/concerns.**
>
> **Main comments:**
>
> **1. Empirical support for overall conclusions: Overall, the conclusions are well-supported by the evidence and the results are reported clearly. A few specific comments:**
>
> **1a. In general, I think that you could do a little more to explain in plain language what each result means and what the broader implication is for people's donation preferences. This would be helpful especially in the abstract, but also in the individual results sections. It can be hard to make sense of everything just based on a list of statistical results.**

Thank you for the feedback. We revised our results section, aiming to make things clearer and in plain language tie the analyses and results to the hypotheses. We made some revisions to the abstract but there is less that we could do there, given word constraints and the need to keep things brief and concise. Please see our reply to minor points 1 and 2.

We then also have the discussion section that ties it all together, and we further ensured a plain language explanation of each hypothesis' results for hypotheses that did not already have one.

For Hypotheses 1 and 2:

> For Hypotheses 1 and 2, our replication results were highly consistent with those of the original study for all of the waste and the past costs conditions, showing that with efficiency kept constant (number of lives saved), people have a clear preference towards charities with lower perceived waste (advertising and overhead) and lower past costs (e.g., setup costs). The effects we found for these two hypotheses were very consistent with the ones provided in the target article, given the change in time, population, and sample size. [...]

For Hypothesis 3:

> [...] we found no support for an effect using a .005 alpha in the control condition in both Study 1 and 2, in line with the results of the original study. These results show support for the hypothesis that not all people allocate all funds to the more efficient charity and instead that on average people tend to diversify their donations to allocate some money to the less efficient charity.

And for Hypothesis 4:

> For Hypothesis 4 (nationalism/ingroup effect), the results support the hypothesis that people tend to favor allocating more money to charities that offer aid to local communities over foreign ones. Unlike the original article which found that "the regions did not differ in allocation", [...]

> **1b. In your waste/overhead replication, it looks like participants allocated much more fairly in study 3 than in studies 1/2. I wonder if you have any ideas why this is. Is it the 0-100 scale vs. the 1-5 scale that matters? Or do you think that specifying how much money remains after overhead in study 3 reduced the size of the effect? Both seem plausible to me.**

That is an interesting observation.

When looking at the charts, the difference in effect size is already apparent. It seems that the most people chose 100% to B in Studies 1 and 2, whereas most people chose 3 (equal allocation) in Study 3.

Our hunch is that it is related to a third reason: the type of waste used in the scenario. Studies 1 and 2 used "advertising", whereas Study 3 keeps it vague as "overhead expenses".

To this effect, we added a section in the main manuscript's Discussion section reading:

> For Hypothesis 1 in particular, the difference between the effect sizes found in Studies 1 and 2 ($d = 0.86$ and 0.84 respectively) and Study 3 ($d = 0.41$) may be due to a difference in wording: "advertising" versus "overhead costs", where "advertising" would have a stronger connotation of waste.

> **1c. In the diversification/unequal efficiency scenarios of studies 1 and 3, I'm not sure it is appropriate to label charity B as "less efficient" or "less effective." Despite having lower EV than charity A, the relative value of these depends on participants' risk preferences. Spreading resources across 5 projects x 70% chance of success per project might be objectively less risky than concentrating resources in one project x 75% chance of success. Perhaps instead of "less effective" you might consider saying something more specific, like "lower EV." (Presumably the original article also made this mistake, but it could still be worth clarifying.)**

We think that this disparity may be an issue that is caused by a difference in definitions. The word "efficient" with regard to donations has a specific definition in the original article, specifically the amount of good that the donation does per sum of money donated. We added this definition to the main manuscript; please see the response to Minor point 4 regarding the definition of "efficiency".

However, we do understand that the word "effective" here can be misunderstood and does indeed hit on your point about a confound with risk aversion, so we changed instances of the word "effective" to "efficient" where applicable for the descriptions of Hypothesis 3 throughout the article:

> To test the diversification effect on charitable giving (Hypothesis 3), participants in Studies 2 and 4 were asked how much they would allocate to a charity that is more efficient compared to a charity that is less efficient… [further changes from "effective" to "efficient"]

2. **Deviations from approved protocol:**

The authors note that their original power analyses were conducted with an alpha level of .05, despite preregistering a threshold of .005. They conclude that "However, given that our SESOI of d = 0.2 was an extreme under-estimation, this did not seem to affect our ability to detect effects." I am not at all concerned that you were underpowered, but I might suggest conducting a post-hoc power analysis anyway to back up this statement (e.g., with an alpha level of .005, what was your power to detect the same effect size? and/or what sample size would you have needed to reach the same level of power?).

Thank you for the suggestion. We agree that this would be a helpful addition.
We have followed this recommendation and added a sensitivity analysis:

> After data collection in Stage 2, we noticed an oversight. We initially conducted our power analysis based on an alpha of .05, and based on the Stage 1 peer review recommendation we adjusted our alpha to .005, yet did not update our power analysis. This did not have much impact, as we targeted d = 0.2 which was an extreme under-estimation. A sensitivity analyses of a sample size of 350 per study (1403/4), power of 0.95, and an alpha of .005, one-sided, showed we were powered to detect a one-sample effect of d = 0.23, and a paired-samples effect of dz = 0.23. We provided more information regarding these calculations in the section on "Analysis of the original article" in the supplementary materials.

and updated the supplementary with the respective G*Power screenshots.

3. **Data quality checks:**

Related to my comment in Stage 1 about potential scale order concerns: I would suggest explicitly stating that the original study didn't randomize the scale order and that this is why you do not do so. To rule out concerns about order effects, something else you could consider reporting in the supplement is a summary of which effects are directionally consistent with a bias toward the right side of the scale (e.g., waste/overhead, where the lower-waste charity is listed second) vs. which are not (e.g., nationalism, where the local charity is listed first). To be clear, I do not believe that this explains your results, but some readers may share my original concern and so I think it would help to show that you've thought about it.

Thank you for the suggestion. We agree that this would be a good addition to the manuscript. Our main takeaway from your very helpful review in Stage 1 is the lack of significant difference between the four scenarios, which is why we focused on that in the supplementary.

Therefore, we added the underlined section below to the main manuscript:

> [...] participants of each study were provided with descriptions of two different conditions, and asked to evaluate the two conditions using various scales; <u>the order of these scales was not randomized, as the original studies did not do so</u>. They were then given an optional open-ended feedback question [...]

and the underlined section below to the supplementary:

> [...] The results showed that there was about a 1-5% difference in allocation percentage based on order in each scenario, <u>with two of the scenarios showing a slight left-hand bias, and two showing a slight bias to the right</u>. However, none of these differences were statistically significant. [...]

4. Confirmatory vs. exploratory analyses:

4a. The authors do a good job of clearly distinguishing which measures/analyses were primary vs. secondary. For example, they chose to include several measures that were included but not reported in the original article, reporting these in supplementary materials. And aside from the potential confusion of using the term "overhead" for two different effects, one from the original article and one extension, they clearly distinguish which research questions fall under each category.

4b. In the results section, I was unsure whether the additional analyses you report (e.g., the one-way ANOVA testing for differences between locations in the nationalism study 3) were exploratory not. Perhaps specifying within each results section which are which could be helpful.

Our original intention was to structure the main manuscript such that all replications of the original analyses came first, followed by the extensions analyses, and finally any exploratory analyses that we conducted post-hoc. All the exploratory analyses that we conducted are under the "Post-hoc exploratory analyses" heading.

To avoid confusion, we renamed the "Replication" and "Extensions" headings of the Results section as "Replication of original analyses" and "Extensions analyses", which hopefully makes the structure of the Results section clearer to the reader.

Minor comments:

.1. PCIRR recommends that reviewers check whether authors have included a direct URL to the approved protocol in the Stage 2 manuscript. Just confirming that the link does appear in the manuscript (on page 17).

.2. In the abstract, and anywhere else you summarize the general pattern of results, I would suggest being clearer about what each result means (e.g., "people preferred to donate to charities with lower perceived waste (stats), lower past costs (stats), …"), as opposed to simply listing the shorthand names for each result as you currently do.

Thank you for this comment. We agree that the summaries can be more explicit.

We have therefore made the following underlined adjustments in the abstract:

> We found support for the effects of a preference for <u>lower perceived waste</u> (d = 0.70, 95% CI [0.41, 0.99]), <u>lower past costs</u> (d = 0.59, 95% CI [0.16, 1.02]), <u>for the ingroup</u> (d = 0.52, 95% CI [0.47, 0.58]), <u>for having some diversification between charities</u> (d = 0.63, 95% CI [0.47, 0.78] for single projects; d = 1.18, 95% CI [1.00, 1.36] for several projects versus one), <u>and against forced charity</u> (d = 0.29, 95% CI [0.21, 0.37]; nominally replicated, but has caveats regarding validity); as at least four of our five hypotheses were found to replicate, we conclude this as being a successful replication

and in the Conclusion section:

> In a very close replication of Baron and Szymanska (2011), we found support for the <u>effects of a preference for lower perceived waste, lower past costs, for the ingroup, and for having some diversification between charities. We also found some indication for preference against forced charity on cost-effective donations, yet with some caveats</u>. [...]

.3. In the study design table, it could be helpful to specify the direction of the "average benefit per dollar" effect. I think lower past costs = higher average benefit per dollar but this is unclear.

Thank you for the suggestion. We added the word "higher" to all instances of Hypothesis 2:

> People prefer to donate to charities with lower past costs (<u>higher</u> average benefit per dollar), even when past costs are irrelevant in the context (H2)

> **.4. In some places (e.g., on page 6), you use the word "efficiency" to refer to (presumably) the overall utility of donating. I would suggest not using this word, or perhaps clarifying how you define it early in the manuscript, because my understanding of the "efficiency" of a donation is something closer to the proportion of funds going to programs (vs. overhead).**

Thank you for pointing this out. We did not consider that the word "efficiency" may mean something else to the reader. We had an implicit definition of how the target article (and by extension we) defined efficiency in the Background section, which we have now made explicit with the underlined section below, adapted from the original article:

> Baron and Szymanska (2011) proposed that utilitarianism, which they defined as "the totality of good that comes about from a choice", should be the objective standard from which to evaluate the efficiency of any given donation. <u>The efficiency of a donation, therefore, is defined as the amount of good that the donation does per unit of money.</u> They argued that charitable donations should be made aiming to maximize the most good possible using the same amount of money.

> **.5. I find it confusing that you use the term "overhead" to refer to both (1) your replication of Baron & Szymanska (2011)'s finding that people prefer to donate to charities with lower overhead and (2) your extension testing whether people prefer to donate when overhead is covered by someone else.**
>
> **Both of these effects would reflect overhead aversion, but to avoid confusion I would suggest labeling the latter just "external funding effect" (i.e., removing "overhead" from the name).**

Thank you for the suggestion. We changed all instances of the title of Hypothesis 7 to "<u>External funding</u>".

> **.6. In table 1, I would suggest replacing the NAs with, for example, hyphens or just leaving these cells blank. The NAs make it difficult to read at a glance.**

We replaced the NAs in Table 1 with hyphens.

> **.7. Thanks for mentioning that your sample size is larger than Simonsohn's recommendation of 2.5x the original.**
>
> **.8. The "measures" starting on page 21 is very clear and helpful. I appreciate seeing the exact wording of each condition.**
>
> **.9. In the diversification with unequal efficiency studies, I would suggest clarifying the logic of how study 2 version 1 tests for a preference for diversification. I realized only after rereading a few times that you are not comparing to the midpoint in this study, but instead comparing to all donations going to the more efficient charity.**

This study is indeed confusing as the unequal efficiency studies have a different *t*-test midpoint compared to the other studies.

For the avoidance of confusion, we have added the following text to the first instance where this occurs in the results section:

> Unlike in the tests of the previous hypotheses, where both of the studies were equally efficient, in the empirical test of this hypothesis, Charity A was clearly more efficient than Charity B in the unequal efficiency scenarios, so the test we conducted was a one-sample t-tests against 0, which allocates all funds to to Charity A.

Reply to Reviewer #2: Dr./Prof. Jonathan Berman

> **The authors conduct a series of replications regarding Baron & Szymanska (2011), with a few extensions. They by and large replicate the work of Baron & Szymanska (2011). Interestingly, they find no evidence that people prefer public over anonymous donations.**

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

> **.1. Re-reading the scenarios in detail, I have a few concerns regarding the wording, and whether participants are fully comprehending them as the researchers intend. For instance, I worry that in the waste/overhead scenario, it may be that some participants think that you need $1,000 per non-overhead money to save 5 lives. If they are thinking this, then they would be right to choose B over A. One way to assess this is to condition Study 3 only on people saving that the "right" allocation is equal?**
>
> **.2. Similarly, in the diversification studies, you can condition results who report the "correct" answer for the impact and/or efficiency questions and then evaluate just those choices. This way, these impact questions would act as a comprehension check and you see whether it is that (a) people just don't comprehend impact appropriately / people are confused by the scenario or (b) even when people realize the impact consequences, they still don't donate in accordance with impact (cf. Berman et al., 2018; Caviola et al., 2020). I wonder if it is worth conducting these additional analyses and commenting on them if necessary.**

Thank you for sharing your thoughts.

We agree that with fictional scenarios, there is always the worry that participants may perceive the questions posed to them differently from how we expect them to see it. The several versions of the scenarios with different framing in the different studies showing similar effects partially gets at that, which is another reason as to why it was valuable to run all these studies together in the same data collection, so that we can see the similar patterns across the different framing.

In addition, the original study had an exploratory qualitative response text box for each item that we included in our replication, so we can examine whether the participants reflect about it in the same way that we expect them to based on their quantitative responses. We use these responses in this way when discussing the results for Hypotheses 5 through 7.

Taking the waste/overhead hypothesis as an example, here are some randomly selected responses from participants from Study 1 and 2:

> *"B seems less wasteful"*
> *"While I understand overhead cost for charities, I have a real hard time when some charities have a HIGH overhead cost. So, I'd feel better with B more."*
> *"They both seem to be as effective, so the amount spent in advertising doesn't bother me."*
> *"Prefer the money to go to the people..."*
> *"Costs vary and depending on how hard it is to generate donations, spending more may be necessary in order to fulfill the needed donations"*

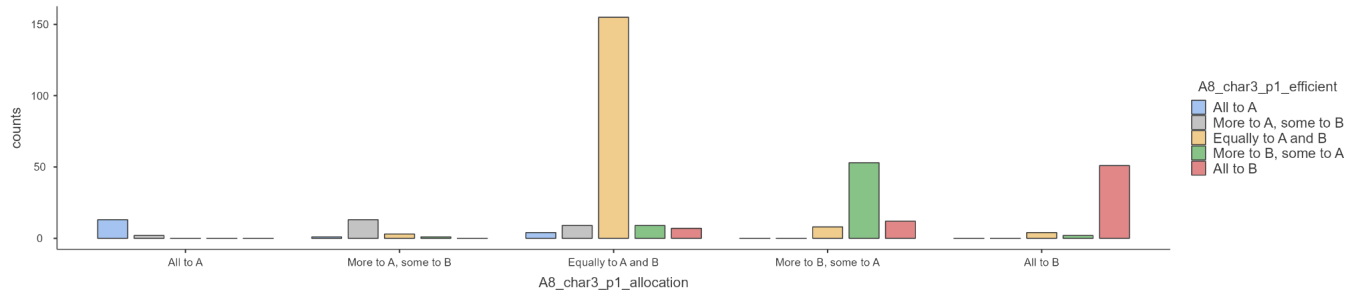
and here are some randomly selected responses from the Study 2 participants on unequal efficiency (version 1):

> *"I can financially save more lives with donating to A."*
> *"Provides the same services for less."*
> *"Although I'm not sure how these funds are to be dispersed, I think it's important to spend money wisely so that more people can be helped."*
> *"A is more cost effective for saving lives."*
> *"From the description, it would seem Charity A is more effective, so I'd rather my money went to them."*

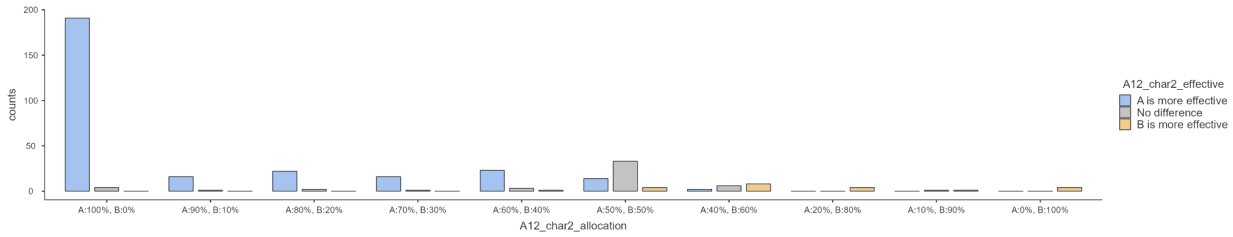
Our impression of the qualitative dataset is that the participants generally perceived the items in the intended ways. Of course we have no way of knowing if the participants who did not give a qualitative response misunderstood the question or not, but from the qualitative data we do not find reason to doubt the original assumption that the items are overall understood in the intended way by the participants, especially considering that if a participant is confused by an item they very well be motivated to say so given a feedback box.

To examine the possibility that participants do not donate in accordance with what they think is the most efficient way, we ran a few exploratory analyses. We found that generally people chose to donate what they believed was the most efficient way to donate.

For example, in Study 3's question on waste/overhead, we found support for perceived efficiency predicting preferred allocation ($F(1,345) = 873$, $p < .001$, $R^2 = 0.717$), with similar effects found for Studies 1 and 2. This implies that what participants think is the most efficient option generally leads participants to select it when allocating, as can be seen in the following plot of allocated donations against perceptions of the most efficient distribution:



We observe a similar pattern with the unequal efficiency items, when the most efficient selection is not a 50%/50% distribution: in Study 2's unequal efficiency items (version 1), we found support for perceived efficiency predicting preferred allocation ($F(1,355) = 405$, $p < .001$, $R^2 = 0.533$):



Overall, it seems that people choose the allocations they chose not because they may have misinterpreted the item or because they chose against what they think is the "correct" distribution, but because they think that what they are doing is indeed the "correct" answer. We provide our datasets and all analyses and code for any future studies to analyze and build on.

> **3. I am suspicious of the external validity of the anonymous donation results. Other work in the marketplace finds an effect of public rewards on donation behavior (e.g., Lacetera & Macis, 2010). One possibility is that social rewards motivates people who would otherwise not donate to charity. That is, people who previously wouldn't donate become motivated to when reputation is on the line. Additionally, it may be that people who have already agreed to donate may subsequentially pass up the opportunity to go public to signal their motives were pure (c.f., Kirgios et al., 2020). These possibilities are worth discussing.**

Thank you for these insights. We are also unsure regarding the external validity of this finding given the literature. There are many factors that may be at play here, which could motivate future research.

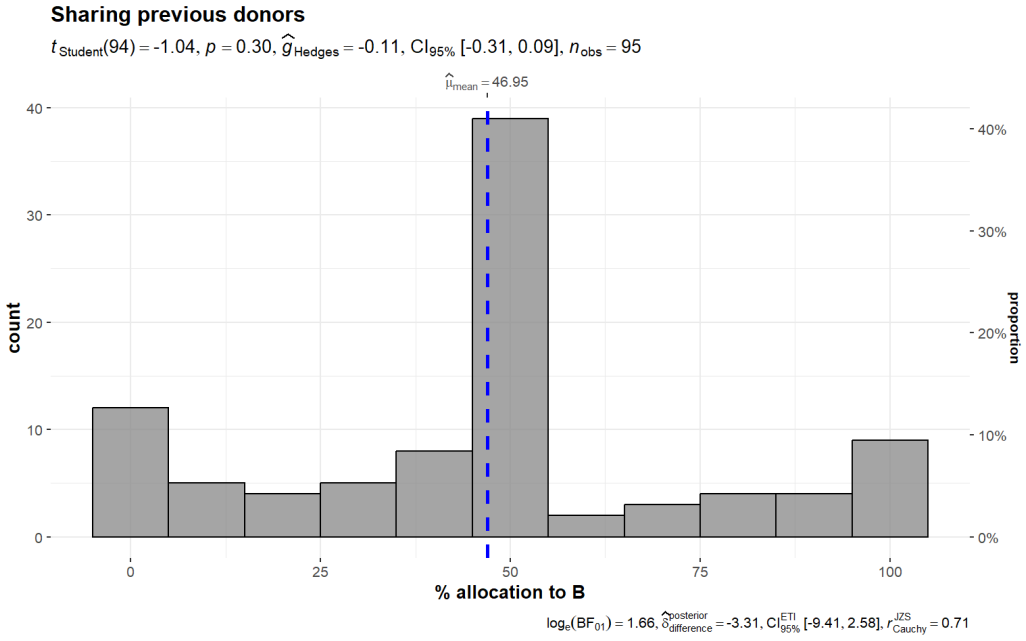
There are experiments such as Vesely et al. (2022) who gave participants money to donate or take home and found that people preferred publicity over anonymity when decisions were

observable. It is possible that people who indicate a preference for anonymity will act differently in real life situations, given that hypothetical scenarios have no real stake.

In a follow-up crowdsourced project of this study, we ran a related scenario:

> Charities A and B both save lives with the same effectiveness (lives saved per dollar spent). Charity A provides information regarding the typical donation amounts by previous donors (without further details), and will include your donation information in future donations. Charity B does not provide any information regarding past donations.
>
> H75a: No sharing > Sharing (>50) (People prefer anonymity) H75b: Sharing > No sharing (<50) (People prefer transparency)

And found no indication for a preference:



So, this offers a bit more nuance that the issue might not be the sharing of information regarding something like donation amount (as in this extension), but rather the specific use of the names of the donors. There is potential for more work even when using hypothetical scenarios.

Therefore, we have made the following changes to the Hypothesis 6 discussion:

> ~~We are uncertain as to why the opposite effect was found in our study,~~
>
> There are a few reasons that this result may have occurred. One is that the results are due to signaling: when given a choice between anonymity and going public, people may choose anonymity over going public to signal that their motives are pure and that they do

> not wish to donate to increase their reputation. Another reason involves the ecological validity of this study: as the scenarios are merely hypothetical and do not involve real money, it may be that in actual scenarios, people may want to get their "money's worth" back from donation and "buy" some reputation using the money that they donated, and signaling becomes less important in priority. We consider this a promising direction for future research.

We also added the following encouraging future studies with real stakes:

> However, we also believe that our methodology can easily be adapted by future research with minimal change needed to investigate between-subject manipulations and other different stimuli as well. New studies may also introduce actual donations with real stakes and compare those to hypothetical scenarios (e.g., Vesely et al., 2022).

> **Minor comments:**
>
> **Page 8. The Caviola cite does not show this exactly. Rather, it says that people weight overhead over effectiveness in between subject design, but not within subjects designs**

Thank you, we revised to the following:

> Recent followup research by Caviola et al. (2014) showed further support for the idea that, when presented separately, people seem more willing to donate more to charities with a low overhead ratio regardless of cost-effectiveness, and that this effect disappears when presented together due to a higher evaluability.

> **References:**
Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to effective altruism: The role of subjective preferences in charitable giving. Psychological science, 29(5), 834-844.
Caviola, L., Schubert, S., & Nemirow, J. (2020). The many obstacles to effective giving. Judgment and Decision Making, 15(2), 159-172.
Kirgios, E. L., Chang, E. H., Levine, E. E., Milkman, K. L., & Kessler, J. B. (2020). Forgoing earned incentives to signal pure motives. Proceedings of the National Academy of Sciences, 117(29), 16891-16897.
Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. Journal of Economic Behavior & Organization, 76(2), 225-237.