



# Understanding biases and heuristics in charity donations

A recommendation by **Romain Espinosa**  based on peer reviews by **Amanda Geiser** and **Jonathan Berman** of the STAGE 2 REPORT:

Mannix Chan, Gilad Feldman (2024) Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011). OSF, ver. 5, peer-reviewed and recommended by Peer Community in Registered Reports. <https://osf.io/4etkp>

Submitted: 27 April 2024, Recommended: 25 September 2024

#### Cite this recommendation as:

Espinosa, R. (2024) Understanding biases and heuristics in charity donations. *Peer Community in Registered Reports*, 100775. [10.24072/pci.rr.100775](https://doi.org/10.24072/pci.rr.100775)

Published: 25 September 2024

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Decisions to give to charities are affected by numerous external and internal factors. Understanding the elements influencing donation decisions is of first-order importance for science and society. On the scientific side, understanding the determinants of charity-giving contributes to the knowledge of altruistic behaviors in the presence of collective problems such as poverty, climate change, or animal welfare. On the social side, pointing out which factors affect donations can help increase prosocial behaviors and might facilitate collective actions in the case of public goods.

Previous work has identified multiple mechanisms affecting altruistic donations to charities (Bekkers and Wiepking, 2011). Importantly, Baron and Szymanska (2011) collected empirical evidence suggesting that people prefer (i) their donations to be directly used for projects rather than organizational costs, (ii) when charities have low past costs, (iii) to diversify their donations into several NGOs, (iv) to favor charities that deal with close peers like nationals, and (v) to give voluntarily rather than through taxes.

Here, Chan and Feldman (2024) conducted a close replication of Studies 1 to 4 of Baron and Szymanska (2011) using a large sample of online participants (four studies, overall N=1,403). In their replication, the authors found supporting evidence for the phenomena reported in the original study. In particular, people were more likely to donate to charities with lower organizational and lower past costs, to diversify their donations, and to show ingroup/nationalist preferences with larger donations to NGOs helping local over foreign children. Chan and Feldman (2024) ran additional analyses that indicated validity concerns regarding the analysis and questions that resulted in finding a preference for voluntary donations over taxation. In their added extensions that went beyond the original study, they also found that donors preferred to donate to charities whose overhead costs are paid for by other donors and unexpected evidence that making donations anonymous increased rather

than decreased contributions. The Stage 2 manuscript was evaluated over one round of in-depth review by the recommender and two expert reviewers. Following revision, the recommender judged that the manuscript met the Stage 2 criteria and awarded a positive recommendation. **URL to the preregistered Stage 1 protocol:** <https://osf.io/gmswz> **Level of bias control achieved:** Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.* **List of eligible PCI RR-friendly journals:**

- [Collabra: Psychology](#)
- [F1000Research](#)
- [International Review of Social Psychology](#)
- [Meta-Psychology](#)
- [Peer Community Journal](#)
- [PeerJ](#)
- [Royal Society Open Science](#)
- [Social Psychological Bulletin](#)
- [Studia Psychologica](#)
- [Swiss Psychology Open](#)

#### **References:**

1. Baron, J. and Szymanska, E. (2011). Heuristics and Biases in Charity. In D. M. Oppenheimer and C. Y. Olivola (Eds.), *The Science of Giving: Experimental Approaches to the Study of Charity* (pp. 215–235). Psychology Press. <https://doi.org/10.4324/9780203865972>
2. Bekkers, R. and Wiepking, P. (2011). A Literature Review of Empirical Studies of Philanthropy. *Nonprofit and Voluntary Sector Quarterly*, 40, 924–973. <https://doi.org/10.1177/0899764010380927>
3. Chan, M. and Feldman, G. (2024). Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011) [Stage 2]. Acceptance of Version 5 by Peer Community in Registered Reports. <https://osf.io/4etkp>

## **Reviews**

### **Evaluation round #1**

DOI or URL of the preprint: <https://osf.io/5vk2m>

Version of the preprint: 4

### **Authors' reply, 07 September 2024**

Revised manuscript: <https://osf.io/4etkp>

All revised materials uploaded to: <https://osf.io/bep78/>, updated manuscript under sub-directory "PCIRR Stage 2\PCI-RR submission following R&R"

**[Download author's reply](#)**

**[Download tracked changes file](#)**

## Decision by [Romain Espinosa](#) , posted 08 July 2024, validated 09 July 2024

### Revision requested

Dear authors,

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. 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.

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.

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.

Best regards,

Romain

### Reviewed by [Amanda Geiser](#), 23 June 2024

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. I will list my main comments first, which are based on the factors that PCI-RR 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:
  - (a) 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.

- (b) 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.
  - (c) 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.)
2. Deviations from approved protocol:
- (a) 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?).
3. Data quality checks:
- (a) 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.
4. Confirmatory vs. exploratory analyses:
- (a) 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.
  - (b) 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.

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.
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.
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).
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.
6. 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).  
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.
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.

Thanks again, and I look forward to seeing this work published!

Amanda Geiser

### **Reviewed by Jonathan Berman, 07 June 2024**

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.

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? 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.

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 subsequently pass up the opportunity to go public to signal their motives were pure (c.f., Kirgios et al., 2020). These possibilities are worth discussing.

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

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.

-Jonathan Berman