

Thank you for the opportunity to review this stage 1 registered report titled “Do pain and effort increase prosocial contributions?: Revisiting the Martyrdom Effect with a Replication and extensions Registered Report of Olivola and Shafir (2013)”.

This pre-registered report is a planned replication of a classical work on the Martyrdom Effect from Olivola and Shafir (2013). Specifically, the researchers aim to replicate and extend findings of study 3, 4 and 5 from Olivola and Shafir (2013). For this replication, they will test the hypotheses proposed in the original article and plan to investigate whether the prospect of enduring pain and exerting effort for a prosocial cause promotes charitable giving. They are closely following the target’s article design while combining study 4 and study 5’s design and collecting data for this combined study and study 3 in one session. Overall, the report is very well-written, easy to follow, and well-structured. It transparently describes the methods, the procedure, and the analytic strategy. Still, I discovered some small inconsistencies and a lack of clarity which I will describe below.

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

The researchers aim to replicate and extend findings of Olivola and Shafir’s (2013) main studies 3, 4 and 5 on the Martyrdom Effect and test an alternative account to explain findings attributed to the Martyrdom Effect. Specifically, they investigate how the willingness to participate, donate and perceived meaning is impacted by effort and pain involved in fundraising activities and secondly how the cause and fundraising type (effortful vs. easy) interact in impacting the willingness to participate. Based on the original article and the outline described in the registered report the research question is considered to be scientifically justifiable.

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

In line with the original article, the authors propose and test hypotheses with regards to the Martyrdom Effect. They test whether greater anticipated effort and costs is correlated with greater prosociality (e.g., donations / willingness to participate in fundraiser activities) and further investigate a moderating effect of cause. Additionally, they aim to test one hypothesis based on the attribute substitution strategy (Kahneman & Frederick, 2002) as an alternative theoretical account that could explain the empirical findings described as evidence for the Martyrdom Effect. It is appreciated, that the authors reframed the original hypothesis for this part of the project to avoid relying on conventional null hypothesis significance testing. To my understanding, the hypotheses are coherent with the theoretical reasoning and are described clearly. However, I would ask the authors to state more explicitly in the main text of the manuscript that the hypotheses described in the original article are the exact hypotheses tested in this replication. This could for instance be added in the section in which the authors describe the hypothesis from the original article (p. 12).

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

Overall, the planned study design is a close replication of the original studies with minor deviations (e.g., combining study design of 4 and 5 and having one unified data collection for

study 3, 4 and 5). The authors sufficiently describe their reasoning behind these small adjustments. The suggested randomization and counterbalancing and a potential analysis of order effects seems reasonable. It is also appreciated that the authors added Manipulation Checks regarding the importance and the effort of the charitable cause which enables a check for internal validity. I would suggest stating explicitly in section Manipulation Checks (extension) that the authors of the original study did not include any Manipulation Checks. The authors based their sampling plan on a power analysis conducted with GPower. While the authors describe some parameters of their power analysis, it would be helpful if the authors could provide more information regarding the type of statistical test and design they used when running the power analysis to increase transparency. The authors state that to detect the smallest effect size for impact of effort on donations, $N = 113$ participants are needed per condition. They then multiply this number by 9 (= number of conditions) which results in a sample of $N = 1017$. Finally, they state that to account for possible exclusions of 10% based on their previous experience and their integrated design they will collect data from 1350 participants. The final sample size is therefore roughly 30% higher than the sample size based on the smallest effect of interest from the original study. However, in the section outliers and exclusion (p. 34) the authors explicitly state that they will include all data of those who successfully completed the study (i.e., they specify no exclusion criteria). Therefore, I would like the authors to clarify the required sample size of 1350 participants. Additionally, the authors mention a “small-telescope” approach in table 6 to explain their power analysis. This approach is not mentioned and explained in the main section of the manuscript. I would suggest to also explain this in the main section of the manuscript.

[Note: The authors mentioned that sometimes their calculations of their effects deviated from the effects mentioned in the original study. Unfortunately, I was not able to identify why these effects deviate.]

The analysis plan proposed by the authors closely follows the approach of the original article. The authors describe which type of statistical analysis will be used to test the respective hypothesis. I would kindly ask the authors to provide more details on the planned statistical analysis e.g., specifying the respective dependent and independent variables and providing information regarding which result would be considered as a finding in line with / or against the proposed hypotheses.

[Note: The authors refer to the RMarkdown for a summary of tests used to test the hypotheses. However, I would kindly ask the authors to already specify the planned test for the hypothesis in the main text of the manuscript (e.g., to test H_x , we conduct a XY with Y as a dependent variable and X as an independent variable).]

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

Overall, the authors provide sufficient methodological detail to be reproducible and ensure protection against research bias. The method section is clearly structured (e.g., through the structuring with headlines and the provision of complex information in tables) and accessible to readers. However, as I already outlined below, the authors should provide more information about the parameters of the power analysis and the planned statistical analysis to test their hypotheses. In the section procedure, the authors describe that “three of four

questions also served as attention checks”. I would kindly ask the researchers to state which questions were considered as an attention check and specify how these attention checks will be used (e.g., participants then cannot take part in the study if they fail the attention checks?).

The researchers also provided us with a Qualtrics link to test the survey. Overall, the survey looks very good. However, I noticed that one had to indicate the donation / pricing for all 20 conditions of distance. I would kindly ask the authors to clarify whether this was an issue in the programming of the study or whether the factor distance is a within-subject factor and not a between-subject factor as indicated in the design of the study.

1E. 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).

The authors have considered sufficient outcome-neutral conditions by including both positive controls and other quality checks. Namely, they added two manipulation checks for the combined studies 4 and 5 to test whether the manipulations indeed have effects on effort and cause importance. Additionally, the authors added attentiveness checks in the beginning of the study.

I hope the authors consider my comments to be helpful in preparing the final version of the registered report.

Best regards,

V. Clemens

Vanessa Clemens