UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Maier, M., Wong, Y. C., & Feldman, G. (2024). Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007). *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.90203

Social Psychology

Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)

Maximilian Maier[1][a], Yik Chun Wong[2][b], Gilad Feldman[3][c]

[1] Department of Experimental Psychology, University College London, London UK, [2] Department of Psychology, University of Hong Kong, University of Hong Kong, Hong Kong SAR, [3] Department of Psychology, University of Hong Kong, Hong Kong SAR

Keywords: identifiable victim effect, bias, judgment and decision making, deliberative thinking, affective heuristics, robust Bayesian meta-analysis, publication bias

https://doi.org/10.1525/collabra.90203

Collabra: Psychology

Vol. 9, Issue 1, 2023

The identifiable victim effect describes the stronger tendency to help a specific victim than to help a group of unidentified statistical victims. Our reanalysis of a meta-analysis on the effect by Lee and Freely (2016) using robust Bayesian meta-analysis suggested publication bias in the literature and the need to revisit the phenomenon. We conducted a pre-registered far replication and extension of Studies 1 and 3 in Small et al. (2007), a seminal demonstration of the identifiable victim effect, with hypothetical donations. We examined the impact of deliberative thinking on the identifiable victim effect both by directly informing participants of the effect (Study 1) and by providing an identified victim with statistical information (Study 3). We found no empirical support for the identifiable victim effect (η_p^2 = .000, 95% CI [.000, .003]) and subsequently no support for debiasing such a phenomenon (η_p^2 = .001, 95% CI[.000, .012]). These findings suggest that the identifiable victim may be better framed in terms of 'scope-insensitivity'. In other words, rather than providing more to a single identified victim, participants seem to be insensitive to the number of victims affected. However, our study involved only hypothetical donations rather than a real-effort real-donation paradigm as in Small et al. (2007). Therefore, we hope that our results spark motivation for future high-powered replications with real money donations, ideally carried out as registered reports and in collaboration with proponents of the original effect. Materials, data, and code were made available on the OSF: https://osf.io/n4jkh/ .

The identifiable victim effect is the tendency to offer more support to an identifiable individual over a group of unidentified victims who are described using numerical statistics (Jenni & Loewenstein, 1997). This inconsistent valuation results in inefficient resource allocation. Small et al. (2007) showed that the identifiable victim effect could be weakened by deliberative thinking, meaning that being informed of the effect and thinking analytically about one's own altruistic behavior may reduce motivation to offer assistance to a single beneficiary. However, the effect was diminished not because participants gave more to the statistical victims, but because they gave *less* to an identified victim. In Study 1 of Small et al. (2007), participants in an explicit learning condition were taught about the identifiable victim effect before making a donation decision. Participants who were briefed about the phenomenon in this way donated less to an identifiable victim compared to the control group. Study 3 found similar results in a condition where the identified victim was presented together with victim statistics: those in the joint presentation condition donated less than those in the identified victim condition, presumably because it reminded them of the many other victims who would not receive help.

We report a replication of Small et al. (2007), in which we had two major goals. Our first goal was to conduct an independent preregistered well-powered conceptual replication of the classic identifiable victim effect on hypothetical donations. This included two manipulations aimed at debiasing the effect: An explicit learning technique, which consisted of informing people about the effect and an implicit learning technique, which consisted of showing the identifiable victim jointly with victim information. Our second

a Contributed equally, joint first author

b Contributed equally, joint first author

c Corresponding Author:
Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; gfeldman@hku.hk

goal was to examine associations between affective feelings and hypothetical donations examining the impact of identifiability and explicit learning on affective feelings. Additionally, we added an extension examining associations with perceived impact of the donation.

The Identifiable Victim Effect

The phenomenon of disproportional generosity provoked by identifiable in comparison to unidentifiable individuals appears to be supported by substantial empirical research (Bergh & Reinstein, 2021; Caviola et al., 2020; Erlandsson et al., 2014; Friedrich & McGuire, 2010; S. Lee & Feeley, 2016; Loewenstein et al., 2006; Slovic, 2007; Small & Loewenstein, 2003). Kogut and Ritov (2005) demonstrated that this effect was restricted to a single target with their name, age, or face displayed, which resulted in larger hypothetical donations compared to a group of unidentified victims.[1] Other moderators were proposed to account for the identifiable victim effect, including the number of identified or unidentified victims, entitativity, cause of plight, perceived responsibility, emotions displayed by the victim, and sense of belonging (Erlandsson et al., 2015; Ritov & Kogut, 2011; Small & Verrochi, 2009; Smith et al., 2013).

There are several possible explanations for the identifiable victim effect. One explanation is proportion dominance, which refers to the phenomenon that people show higher sensitivity to proportions than to absolute values (Baron, 1997). This heuristic suggests that people pay more attention to proportions or percentages than to absolute numbers. Therefore, when evaluating options to save lives, a higher proportion of lives saved seems to result in more helping (Erlandsson et al., 2014; Jenni & Loewenstein, 1997). In the case of an identifiable victim, given that the victim serves as the only reference, the proportion is perceived as 100%. For statistical victims, on the other hand, the reference group may consist of millions of people. Even though the absolute number of lives saved would be higher than a single individual, the proportion of lives saved decreases, thus reducing willingness to help.

Another explanation for this effect is the 'affect heuristic' (Slovic et al., 2007). This heuristic describes the tendency to rely on emotional and affective states when evaluating a stimulus, and is believed to be activated when evaluating a specific identified victim (S. Lee & Feeley, 2018; Small & Loewenstein, 2003). The personalized information received about a specific victim is argued to elicit stronger affective reactions such as sympathy and distress, which likely motivates the willingness to offer more support to that individual. In contrast, a general number representing statistical unidentified victims may fail to induce any major affective responses, and therefore lead to less willingness for altruistic behavior.

Finally, the effect may also be explained by the perceived impact of the donation (Erlandsson et al., 2015). If participants have a stronger belief in the impact of their donations, they are likely to be more willing to give. Duncan (2004) reported that donations towards identifiable victims were perceived to have stronger impact, likely because it was easier to picture how the money or resources would benefit the individual, compared to statistical victims who were depicted in an abstract manner or as a number. However, findings seem mixed regarding the mediating effect of perceived impact on the identifiable victim effect. For example, Lee and Feeley (2016) and Friedrich and McGuire (2010) did not find support for this factor, reporting no differences in the helping behavior between a personalized individual and an anonymous group of people.

Reanalysis of a Meta-Analysis on the Identifiable Victim Effect Suggests the Need to Revisit the Phenomenon

Lee and Freely (2016) conducted a meta-analysis that summarized 41 effects from 22 experiments on the identifiable victim effect and found a 'significant yet modest IVE [identifiable victim effect]' (S. Lee & Feeley, 2016, p. 199) referring to an aggregated effect of $r = .05$. However, there is reason to believe that this effect might be even weaker when publication bias is accounted for: the three highest powered studies in the dataset show effects that are almost zero, including one study with 12802 participants ($r = 0.004$). Lee and Freely examined the possibility of publication bias using visual inspection of funnel plots. However, this approach does not perform well under some conditions like high heterogeneity (Bartoš, Maier, Quintana, et al., 2022; Bartoš, Maier, Wagenmakers, et al., 2022; Carter et al., 2019; Hong & Reed, 2021; Kvarven et al., 2020; Lau et al., 2006; Maier, VanderWeele, et al., 2022), which is present in Lee and Freely's meta-analysis (Q_T [40] = 104.65, $p < .001$, $I^2 = 61.8\%$).

Many have proposed alternative bias correction techniques (for reviews see Carter et al., 2019; and Renkewitz & Keiner, 2019), but these only perform well under some meta-analytic conditions in terms of effect size, heterogeneity, and publication bias. As it is not possible to know the meta-analytic conditions without having adjusted for publication bias, this situation poses the following Catch-22 problem: In order to adjust for publication bias, one needs to know the data generating process, but in order to know the data generating process, one needs to have adjusted for publication bias (Bartoš, Maier, Shanks, et al., 2022). Robust Bayesian Meta-Analysis (RoBMA) is a novel method that aims to overcome this problem using Bayesian model-averaging (Bartoš, Maier, Wagenmakers, et al., 2022; Bartoš & Maier, 2020; Maier, Bartoš, et al., 2022). Instead of selecting a single model, RoBMA applies mul-

1 Though we note a recent failed replication of the Kogut and Ritov (2005) by Majumder et al. (2023).

tiple models simultaneously and allows the data to guide the inference to be based most strongly on those models that predicted the data best. This multi-model inference avoids the Catch-22 problem discussed above. Specifically, RoBMA includes models of selection for significance (Vevea & Hedges, 1995) and models based on the relationship between effect sizes and standard errors (precision effect test & precision effect estimate with standard errors, PET-PEESE). Rather than selecting a single model, Bayesian model-averaging bases the inference on all models (the two publication bias correction methods above as well as methods assuming no publication bias) and weighs them based on how well they predict the data. Therefore, it is much more robust to model misspecification compared to previous publication bias adjustment methods.

RoBMA outperformed other methods for publication bias correction in a large simulation study (Hong & Reed, 2021, reanalysis with RoBMA in Bartoš, Maier, Wagenmakers, et al., 2022), which combined the simulation environments from four previous studies (Alinaghi & Reed, 2018; Bom & Rachinger, 2019; Carter et al., 2019; Stanley et al., 2017). In addition, RoBMA has also been shown to perform better than other methods on empirical data by comparing the estimates of bias-adjusted meta-analyses to registered replication reports (Bartoš, Maier, Wagenmakers, et al., 2022). Here we use the version of RoBMA (also known as RoBMA-PSMA [publication selection model averaging]) as in Bartoš, Maier, Wagenmakers, et al. (2022), as it has been vetted extensively in simulation studies and applied examples (in the same paper). For details about the 36 models that are included, and the corresponding prior distributions and prior model probabilities see Bartoš, Maier, Wagenmakers, et al. (2022).

RoBMA quantifies evidence using Bayes factors. Bayes factors compare the likelihood of the data under competing models (in our case, the alternative hypothesis in comparison to the null hypothesis). In our paper we report BF_{01}. In other words, Bayes factors have the null in the numerator and the alternative in the denominator, and denote evidence in favor of the null hypothesis. As a rule of thumb for Bayes factors with the null in the numerator, Bayes factors between 1 and 3 are often regarded as weak evidence for the null, Bayes factors between 3 and 10 are often regarded as moderate evidence for the null, and Bayes factors larger than 10 are often regarded as strong evidence for the null (e.g., Jeffreys, 1939; M. D. Lee & Wagenmakers, 2013, p. 105; Wasserman, 2000). However, we caution that these rules of thumb should merely aid interpretation and not be taken as absolute thresholds. Bayes factors are continuous measures of the strength of evidence, and any discretization inevitably results in loss of information.

When applying RoBMA to the data by Lee and Freely (2016), we found moderate evidence for publication bias (BF_{01} = 0.11) and strong evidence for the absence of the average effect (BF_{01} = 14.93), with a model-averaged mean effect size estimate of r = 0.002 (95% CI [0; 0.004]).[2] In addition, we find weak evidence against heterogeneity (BF_{01} = 1.24). We plotted the pattern of bias in Lee and Freely in Figure 1. The left panel shows the regression line of effect sizes on standard errors. This relationship indicates that studies with smaller standard errors show smaller effects, a pattern that is indicative of publication bias. The right panel shows the relative publication probabilities for nonsignificant in comparison to significant p-values. This panel indicates that nonsignificant studies (p > .05) are considerably less likely to be published than significant studies. Note that most of the posterior probability among the publication bias models is on the selection models rather than models assuming a relationship between effect sizes and standard errors (see supplementary materials).

Choice of Target for Replication: Small et al. (2007)

We chose Studies 1 and 3 of Small et al. (2007) for replication due to the article's considerable impact. At the time of writing (April 2023), there were 1210 Google Scholar citations of the target article. Beyond the direct citation count, Small et al. (2007) have influenced several other highly cited articles (> 1000 times at the time of writing; e.g., Bekkers & Wiepking, 2011; Slovic, 2007) and popular science and philosophy books such as 'The Life You Can Save' (Singer, 2019) and 'Poor Economics' (Banerjee & Duflo, 2011), which have guided both research and policy. Furthermore, charities often feature pictures of identified victims in advertisements, hoping to employ this effect to increase charitable giving (e.g., https://www.savethechildren.org.uk/), underscoring the applied importance of Small et al.'s findings.

To our knowledge, there has been one direct replication of Small et al. (2007): a Spanish language unpublished doctoral thesis failed to find support for the results of Study 1 (Charris, 2018). However, Charris (2018) only found weak evidence against the effect in a Bayesian analysis and no evidence for the null using the TOST procedure to test for equivalence (e.g., Lakens et al., 2018). Charris (2018)

2 Due to the lack of publication bias correction methods that can accommodate a three-level structure, we accounted for the dependency by only using the most precise estimate within each experiment. Often there were multiple estimates with the same precision within a study. In this case, we selected randomly and bootstrapped 500 times. Using the median of these bootstraps, this analysis comes to the conclusions regarding evidence for publication bias and evidence for an effect. Unlike the main analysis we find moderate rather than weak evidence against heterogeneity. In addition, as funnel plot based methods are sometimes criticized for finding bias for reasons other than publication bias (Lau et al., 2006; Maier, VanderWeele, et al., 2022), we also reanalysed the meta-analysis using only the selection models in RoBMA. This lead to the same conclusions. As only one of the authors is familiar with RoBMA we also requested an independent verification to double check our analysis, the corresponding r script is available in the supplementary materials. as the analysis including selection models.

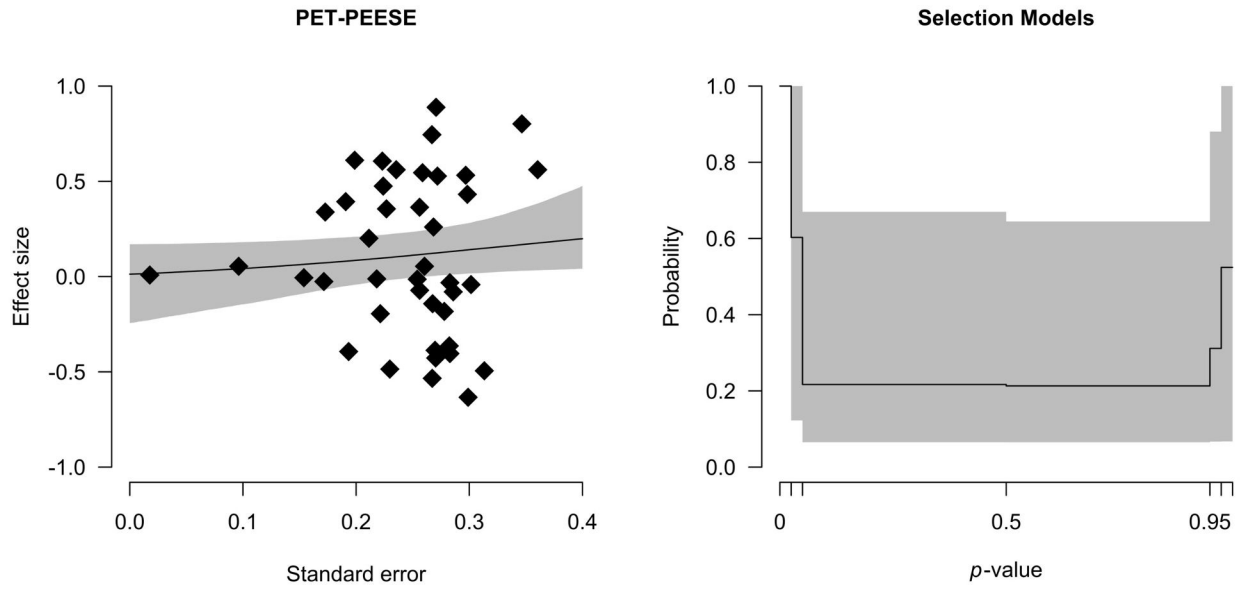


Figure 1. Footprint of Publication Bias in Lee and Freely (2016)

Note. The left panel shows the PET-PEESE regression line (i.e., the relationship between effect sizes and standard errors) and the right panel shows the relative publication probabilities based on the selection models. The left panel displays a regression line of effect sizes on standard errors, the intercept of this line indicates the hypothetical estimate of a study with infinite precision; we can see that it is very close to 0. The right panel displays estimates for the relative publication probabilities of nonsignificant studies in comparison to significant studies model averaged across the different selection models included in RoBMA.

concluded that his study lacked statistical power and does not allow rejecting the identifiable victim effect. In other words, more evidence is needed using high-powered direct replications. Several other recent studies have also questioned the robustness of the phenomenon, but usually only in conceptual replications. For example, Hart, Lane, and Chinn (2018) failed to find support for variations in people's prosocial responsiveness focusing on a single victim than many individuals. Recently, Moche and Västfjäll (2021) and Moche (2022) also failed to replicate the effect across 6 of 7 well-powered studies. A field experiment also failed to provide evidence for the effect (Lesner & Rasmussen, 2014). These failed replications are surprising given that other high-powered studies did find evidence for the identifiable victim effect (e.g., Caviola et al., 2020; Galak et al., 2011; Sudhir et al., 2016).

However, conceptual replications are limited in their ability to inform about previous findings, as when conceptual replications failed it can be argued that the differences in methodology are the explanation for the different results (Chambers, 2017, p. 16). This may interact with a file-drawer and publication bias problems in a literature, that may result in a literature with successful conceptual replications but few shared null results.

The combination of the mixed evidence from replications, the above meta-analysis reanalysis, and the impact of Small et al.'s findings, suggests that more research is needed to revisit and reassess the identifiable victim effect using high-power preregistered replications (Isager et al., 2021). We note that we initially set out to conduct a direct close replication, yet decided on first running a far conceptual replication using the same design with an important adjustment of the dependent variable to use hypothetical donations rather than real donations. We did this for a number of reasons. First, this project was related to a different replication project we conducted in Majumder et al. (2023) in which we failed to replicate the identifiable victim effect demonstrated by Kogut and Ritov (2005) who showed the effect using hypothetical donations, as many other studies examining the identifiable victim effect have. We aimed to make the two replications as similar as possible in their dependent variables to allow one replication to possibly inform the other. Second, we acknowledge the differences between hypothetical and real-life behavior, yet thought it best to ensure that the effect holds with simpler hypothetical donations before embarking on a more complex and costly real donation study. Mean donations are typically higher for hypothetical donations than for real donations (Bekkers, 2006); however, we are not aware of any evidence of mechanisms that result in differences between conditions when switching from real to hypothetical donations.

Given this important adjustment regarding the dependent variable, we categorized this replication as far and conceptual, even though much of the rest of the study remains the same. Thus, we caution against over-interpreting from this replication to the original article's real donations effect replicating, though we hope the community would find this informative in the generalizability of the original's design to hypothetical scenarios. We discuss this point and implications in the general discussion.

Small et al. (2007): Hypotheses and Findings

Small et al. (2007) proposed that thinking analytically about the value of lives reduced giving to an identifiable victim but not to statistical victims. They also suggested that implicitly inducing analytical reasoning about the value of lives reduced donations to an identifiable victim but not to statistical victims. They conducted four experiments, and the current replication focused on Studies 1 and 3.

Study 1 Design and Findings

In Study 1, participants were randomly assigned to one of two conditions, with the intervention group learning about the identifiable victim effect from previous research (explicit learning condition), whereas another served as a control group. They were further randomly assigned to either the statistical victim condition, in which they read information either about the problem of starvation in different African countries, or to the identifiable victim condition, in which they received a brief description of an African girl from the Save the Children website. They were then instructed to donate any five one-dollar bills received earlier from a survey to victims they had read about in the letter. After their donation, participants rated different affective reactions they experienced towards the described victim(s). These items included feeling upset, touched, sympathetic, and morally responsible, as well as the perceived appropriateness of donating to help the described victims.

To summarize, their Study 1 design was a 2 (Identifiability: identifiable vs. statistical) x 2 (Explicit Learning: intervention vs. control) between-subjects factorial design. Their results showed that in the control condition without the intervention, donations to the identifiable victim were higher than donations to statistical victims. However, the pattern was different for the participants who were assigned to the explicit learning intervention conditions and learned of the identifiable victim effect before asking to donate, with the donations being similar towards the identifiable victim compared to towards statistical victims. The explicit learning intervention, therefore, seemed to have eliminated the additional donations given towards an identifiable victim.[3] In addition, they showed that aggregated feelings predicted donation behavior better in the identifiable victim/no intervention condition than in the other conditions.

Study 3 Design and Findings

In Study 3, Small et al. (2007) further studied the effect of implicit learning by adding a third identifiability condition, a joint condition (also referred to as "implicit learning condition") that included both a picture of the single victim and general victim statistics, resulting in a three conditions design (identifiable vs. statistical vs. joint). The donation in this joint condition was intended for the described identified victim. The presentation of victim statistics was meant to implicitly eliminate the identifiable victim effect in the joint condition arguably because providing statistics alongside the victim reminds the potential donor of the many people who would not receive help. Study 3 did not investigate how feelings predicted donations. In summary, the Study 3 design included one factor with three levels/conditions: identifiable victim, statistical victims, and the joint/implicit learning condition.

Small et al. (2007) found support for implicit learning, as donations to the identified victim were lower in the joint condition compared to the identifiable victim condition.

Overview of the Replication and Extension

In this replication, we merged Studies 1 and 3 in Small et al. (2007) into a single experimental design to study both the explicit and implicit ways of debiasing the identifiable victim effect. Our study was a 3 x 2 experimental design varying identifiability of the victim (identifiable victim, statistical victims, and joint - identifiable victim alongside statistical victims) and Explicit Learning (present or not). We summarized the design in Table 1.

The Identifiability factor, therefore, included the implicit learning intervention from Study 3 in Small et al. (2007). Extending the original studies, the explicit learning intervention was also manipulated on the joint condition and participants in the joint condition also rated affective feelings. We note that in-line with Small et al. (2007), the donations in the joint condition went towards the identified victim (rather than the statistical victims that were also described in this condition). Mirroring Small et al. (2007)'s Study 1, we assessed aggregated affective feelings as a predictor of hypothetical donations.

We summarized the hypotheses of the current replication in Table 2. To replicate the results of the original study, our Hypothesis 1 tests the identifiable victim effect based on the contributions toward different victims. We combined the original Hypotheses 1 and 2 stated in Small et al. (2007) into Hypothesis 2 to investigate whether being informed about the identifiable victim effect affected donations towards the different victims. In Hypothesis 3, we explored whether learning about the identifiable victim effect affected donations regardless of Identifiability. Hypothesis 4 describes the main effect of implicit learning (i.e., the joint condition; being presented with victim statistics in the identifiable victim condition) to replicate Study 3 from Small et al. (2007). We proposed Hypothesis 5 to examine the impact of Identifiability and Explicit Learning on affec-

3 We note that this differed from our expectations, given that in the charitable giving literature interventions are typically meant to increase donations, and therefore we had expected that such an intervention would increase donations towards statistical victims to the level of donations towards the identifiable victim.

Table 1. Replication and Extension: Experimental Design

		Identifiability (IV1; between-subject)		
		Identifiable victim condition	Statistical victim condition	Joint condition (Implicit Learning)
Explicit Learning (IV2; between-subject)	Explicit learning intervention	Identifiable Explicit	Statistical Explicit	Joint Explicit
	No intervention (Control)	Identifiable Control	Statistical Control	Joint Control

Note. Joint condition displayed both an identifiable victim and general victim statistics.

tive feelings, and added Hypothesis 6 to test our extension regarding perceived impact of donation. For each of the hypotheses, we had hypotheses serving as replications mirroring the target article's designs (Study 1 without the joint conditions, and Study 3 without the Explicit Learning conditions), and additional extension hypotheses that aim to make the most of the unified design using all relevant conditions (with joint, and with explicit learning intervention).

Extension: Perceived Impact of Donation

Given the conflicting findings regarding the influence of the perceived impact of donations we discussed above (Duncan, 2004; Friedrich & McGuire, 2010), we aimed to extend the replication study by also considering the perceived impact of donations. We, therefore, included an additional measure of perceived impact of the donation to investigate whether people consider donations more impactful towards an identifiable victim or statistical victims.

Preregistration and Open Science

We provided all materials, data, and code at: https://osf.io/n4jkh/. We preregistered the study, and the preregistration can be accessed at: https://osf.io/dc9kb/.

We report all measures, manipulations, and exclusions conducted for this investigation. The study was preregistered with power analyses reported in the supplementary materials, and analyses were only conducted after all data had been collected. Deviations from our preregistration are stated in the 'Deviations from preregistration' section of the supplementary materials and also at the appropriate places in the methods section of the main text of the manuscript.

Method

Participants

The study received ethics approval from the University of Hong Kong (EA1908020). A total of 1004 Amazon Mechanical Turk (MTurk) participants were recruited from a US sample using CloudResearch/TurkPrime (Litman et al., 2017; M_{age} = 39.4, SD = 12.4; 465 females, 533 males, 6 prefer not to say). We compared the target article and the replication samples in Table 3.

We collected as many participants as we could afford with the available funding. The full report of power analysis can be found in the supplementary materials under the section 'Power analysis of the original study effect' and indicates that for the lowest powered effect (the interaction between Explicit Learning and Identifiability), a sample size of 314 would be sufficient to achieve 95% power for the original effect size. In addition, sensitivity power analyses indicate that our sample size would have 95% power to detect a very small effect size of $\eta_p^2 = 0.012$ with an alpha level of .05.

Exclusion Criteria

We pre-registered that "We will focus our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we will also determine further findings reports with exclusions", with several exclusions criteria for the supplementary analyses: low English proficiency (scored lower than 4 on a scale of 0 to 6); not being serious in completing the survey (scored lower than 3 on a scale of 0 to 4); correctly guessed the hypotheses; already seen the survey before; failure to complete the survey or completed in less than a minute; and not from the United States.

Fifty-six responses met the exclusion criteria. We found no major differences between the pre- and post-exclusion results. As preregistered, we focused on the full sample for data analysis. We summarized the results after exclusion in the supplementary materials ('Exclusion based on preregistration criteria'), with a comparison of the findings ('Pre-exclusions versus post-exclusions').

We had preregistered using median absolute deviance (MAD) to detect univariate outliers; however, we realized that this procedure is not relevant to our dataset due to the boundedness of all the scales used.

Design and Procedure

We combined Studies 1 and 3 of Small et al. (2007) into a unified between-subject design with a 3 (Identifiability: identifiable vs. statistical, vs. joint) by 2 (Explicit Learning: intervention vs. control) random-assignment experimental design, and with donations and feelings as the dependent variables. We provided additional details regarding the procedure in the 'Procedure' subsection in the supplementary materials and the Qualtrics survey is provided with the preregistration in the OSF folder.

Table 2. Replication and Extension: Summary of Hypotheses

Hypotheses	Label	Hypothesis description	Conditions comparisons for hypotheses
Donations			
1a (Identifiability main effect, without joint) [S1]	Identifiable victim effect in donations	People donate more when presented with an identifiable victim than when presented with statistical victims	Identifiable (Explicit & Control) > Statistical (Explicit & Control)
1b (Identifiability main effect, with joint) [E] *			Identifiable (Explicit & Control) > Statistical (Explicit & Control) ~= Joint (Explicit & Control)
2a (Interaction effect, without joint) [S1]	Explicit learning reduces identifiable victim effect in donations	The identifiable victim effect is weaker for people who were explicitly informed about the identifiable victim effect.	Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control
2b (Interaction effect, with joint) [E] *			Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control ~= Joint-Explicit minus Joint-Control
3a (Explicit Learning main effect, without joint) [S1] *	Explicit learning reduces donations	People who were explicitly informed about the identifiable victim effect tend to donate less than those uninformed of the effect.	Explicit (Identifiable, Statistical) < Control (Identifiable, Statistical)
3b (Explicit Learning main effect, with joint) [E]			Explicit (Identifiable, Statistical, and Joint) < Control (Identifiable, Statistical, and Joint)
4 (Identifiability with Implicit Learning main effect, without Explicit) [S3] *	Statistical information reduces donations towards identified victim (Implicit learning)	People donate less to an identifiable victim when the identifiable victim is presented alongside information about statistical victims (joint condition)	Identifiable (Control) > Statistical (Control) ~= Joint (Control)
Affective feelings			
5a (Identifiability main effect, without joint) [S1] *	Identifiable victim effect in affective feelings	People rate higher affective feelings towards an identifiable victim than towards statistical victims and to an identifiable victim presented alongside statistical victims	Identifiable (Explicit & Control) > Statistical (Explicit & Control)
5b (Explicit Learning main effect, without joint) [S1] *	Explicit learning reduces affective feelings	People who were explicitly informed about the identifiable victim effect tend to donate less than those uninformed of the effect.	Explicit (Identifiable, Statistical) < Control (Identifiable, Statistical)
5c (Interaction effect, without joint) [S1] *	Explicit learning reduces identifiable victim effect in affective feelings	The identifiable victim effect in affective feelings is weaker for people who were explicitly informed about the identifiable victim effect	Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control
5d (Interaction effect, with joint) [E]			Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control ~= Joint-Explicit minus Joint-Control
Perceived impact			
6 (Interaction effect, with joint) [E]	Identifiable victim effect in perceived impact	People rate higher impact for donations to an identifiable victim than towards 1) statistical victims and 2) an identifiable victim presented with victim statistics	Identifiable (Explicit & Control) > Statistical (Explicit & Control) ~= Joint (Explicit & Control)

Note. For the interaction effects, we visually examined that the effect is in the correct direction if the interaction test is significant.
The preregistration only specified the tests including the joint condition (i.e., did not specify H1 & H2). However, we added these hypotheses to ensure a fair comparison to the original article. Donations mentioned in the hypotheses refer to hypothetical donations. [S1] mirrors the target article's Study 1. [S3] mirrors the target article's Study 3. [E] indicates an extension. * indicates analysis was not pre-registered and added for completeness of reporting addressing peer review.

Manipulations

Explicit Learning

Participants were randomly assigned into either the explicit learning intervention condition or to the control condition (evenly presented with the Qualtrics randomizer). Participants in the explicit learning intervention condition were instructed to read a passage about prior research findings on the identifiable victim effect used in the original studies. In other words, they were taught about the phenomenon before the donation.

Table 3. Samples: Comparison of Original Study and Replication

Demographics	Small et al. (2007)	Replication
Sample size	Study 1: 121 Study 3: 159	1004
Geographic origin	US American	US American
Gender	Not specified	533 males, 465 females 6 prefer not to say
Median age (years)	Not specified	36
Average age (years)	Not specified	39.4
Age range (years)	Not specified	20-91
Medium (location)	Laboratory	Computer (online)
Compensation	US$5	Around US$1
Year	2007	2020

Identifiability

Participants were then randomly assigned to one of the three Identifiability conditions. Those in the identifiable victim condition read about a child from Zambia suffering from starvation, accompanied by a black-and-white photograph and a short description. Those in the statistical victim condition read about numerical victim statistics to illustrate the millions of people living in a similar plight to the child described in the identifiable victim condition. The joint condition was a combination of the previous two conditions, with the same Zambian child presented in a photo with a brief description, along with the victim statistics provided in the statistical victim condition; the order of the presentation was randomized (evenly presented with the Qualtrics randomizer).

Forced Manipulation Comprehension Checks

To ensure reading and comprehension of the scenarios, we added checks that the participants had to answer correctly in order to be able to proceed to the next page that presented the dependent measures. This is a noted deviation to the target article's design, which we added to address concerns that online sample participants may not have read or were inattentive to the scenario and the manipulation.

Measures

Hypothetical Donation

Participants were then presented with the following continuation of the scenario: "Imagine that you have just earned $5 US dollars and you are given an opportunity to donate any amount of the money to the organization Save the Children". They then indicate their hypothetical donations from 0 to 5 US$ in increments of $1 ($0, $1, $2, $3, $4, or $5). The donation was to the specific victim in the identifiable and joint conditions and to the anonymous group in the statistical victim condition.

Affective Reactions (with Perceived Impact Extension)

Participants indicated their affective reactions at the time of donation on a 5-point Likert scale, ranging from 1 (*Not at all*) to 5 (*Extremely*). The affective measures were: 1) Upset: "How upsetting is the described situation of the victims to you?", 2) Sympathetic: "How sympathetic did you feel while reading the description of the victims?", 3) Responsibility: "How much do you feel it is your moral responsibility to help out the victims?", 4) Touched: "How touched were you by the described situation of the victims?", 5) Appropriateness: "To what extent do you feel that it is appropriate to give money to aid the victims?", and 6) Perceived impact (extension): "How confident were you that donating your money to the described victims could have a significant impact?". In line with Small et al. (2007) we investigated the effect of each feeling individually as well as the effect on aggregated feelings (without the extension perceived impact).

Replication Closeness Evaluation

We summarized our evaluation criteria of the replication closeness based on Lebel et al. (2018) in Table 4, categorizing the replication as 'far', given the adjustment we made to the dependent variable being a hypothetical scenario instead of a behavioral measure examining real-life donations.

Results

We followed and extended the analyses conducted by the target article. We provided a comparison of the statistical tests reported in the original study and the replication in the supplementary materials.

Descriptive Statistics

We summarized the descriptive statistics for hypothetical donations, aggregated feelings, and perceived impact of the donation in Tables 5-7 and statistical tests for hypothetical donations in Tables 8 and 9. We provided the re-

Table 4. Classification of the Replication, Based on Lebel et al. (2018)

Design facet	Replication	Details of deviation
IV operationalization	Same	/
DV operationalization	Different	**Hypothetical donations** See "DV stimuli" below in subsection "Hypothetical imaginary donation" for details
IV stimuli	Similar	**Updated victim information** We presented the participants with the most updated victim information retrieved from the website of Save the Children given that those used in the original study was dated two decades ago. **Explicit learning intervention applied to the joint condition** To combine the original studies 1 and 3 into a single study, we applied the explicit learning intervention to the joint condition as an extension.
DV stimuli	Different	**Question adjustment** We made minor changes to the questions evaluating participants' feelings to ensure the reported emotion status was linked to the victim information they have just read. **Hypothetical imaginary donation** We asked the participants to indicate how much they would hypothetically donate from $0 to $5 to the corresponding victim(s) instead of donating the money to Save the Children after signing a charity request letter. **Feelings variables measurement** We considered affective feelings in the joint condition which was not measured in the original study. We also added an extension of 'perceived impact of donation' into the scale.
Procedural details	Similar	**No reward acquired from a pre-survey** Participants would not conduct an irrelevant survey prior to the experiment to earn $5 for the donation.
Physical settings	Different	**Online survey** Participants conducted an online survey in Qualtrics on the MTurk platform whereas the original study surveyed inside a student center of a university in Pennsylvania.
Contextual variables	Different	**MTurk workers as participants** We recruited participants on the MTurk platform while the original study recruited participants sitting inside the school center of a university in Pennsylvania.
Replication classification	Far replication	

Note. IV= Independent variable. DV= Dependent variable.

Table 5. Hypothetical Donations: Descriptives

	Identifiable victim condition	Statistical victim condition	Joint condition	Total
Explicit learning intervention condition	2.84 [1.89] {1.36}* (170)	2.74 [1.98] {1.26}* (159)	2.23 [1.91] {N/A} (173)	2.60 [1.94] (502)
No intervention condition	2.58 [1.87] {2.83}* (165)	2.72 [1.92] {1.17}* (176)	2.48 [1.99] {1.43}** (161)	2.60 [1.93] (502)
Total	2.71 [1.88] (335)	2.73 [1.95] (335)	2.35 [1.95] (334)	2.60 [1.93] (1004)

Note. Statistics are presented in the following format: mean [standard deviation] {Small et al. (2007)'s reported means} (condition sample size). *Based on Small et al. (2007) Study 1. **Based on Small et al. (2007) Study 3.

sults for the individual measures of feelings in the supplementary materials.

Hypothetical Donations

We plotted hypothetical donations by conditions (including joint condition) in Figure 2. We summarized the inferential tests of our replication in comparison to Small et al. (2007) in Table 8.

Table 6. Aggregated Feelings: Descriptives

	Identifiable victim	Statistical victim	Joint	Total
Explicit learning intervention	3.82 [0.91] (170)	3.82 [1.02] (159)	3.60 [0.97] (173)	3.75 [0.97] (502)
No explicit learning intervention	3.81 [0.90] (165)	3.84 [1.00] (176)	3.77 [1.05] (161)	3.81 [0.98] (502)
Total	3.81 [0.91] (335)	3.83 [1.01] (335)	3.69 [1.01] (334)	3.78[0.98] (1004)

Note. Statistics are presented in the order of Mean [Standard deviation] (condition sample size). We reported the same information for the non-aggregated feelings in the supplementary materials. Aggregated feelings were calculated following the approach by Small et al. (2007): Upset, sympathetic, touched, responsible, and appropriateness. The Cronbach's alpha for the five feelings measures was 0.90.

Table 7. Perceived Impact (Extension): Descriptives

	Identifiable victim	Statistical victim	Joint	Total
Explicit learning intervention	3.47 [1.23] (170)	2.91 [1.37] (159)	2.94 [1.35] (173)	3.11 [1.34] (502)
No explicit learning intervention	3.31 [1.25] (165)	2.99 [1.34] (176)	3.11 [1.42] (161)	3.14 [1.34] (502)
Total	3.39 [1.24] (335)	2.96 [1.35] (335)	3.02 [1.39] (334)	3.12 [1.34] (1004)

Note. Statistics are presented in the order of Mean [Standard deviation] (condition sample size).

H1a, H2a, and H3a: Identifiability and Explicit Learning Main Effects and Interaction (without Joint Condition) [Replication]

Following the analyses conducted in Study 1 of Small et al. (2007),[4] we carried out a 2 (Explicit Learning) × 2 (Identifiability) two-way ANOVA (i.e., cells Identifiable-Explicit, Statistical-Explicit, Identifiable-Control, and Statistical-Control) to examine the following hypotheses: 1) H1a: People donate more when presented with an identifiable victim than when presented with statistical victim, 2) H2a: The identifiable victim effect (H1a) is weaker for people who were explicitly informed about the identifiable victim effect 3) H3a: People that were explicitly informed about the identifiable victim effect tend to donate less than those uninformed about the effect.

We supplemented the frequentist analyses with a Bayesian analysis to allow quantifying evidence for the null. As parameter prior distribution we use a Cauchy (0, 0.707), which is a common choice in Bayesian analysis. Because the Cauchy distribution is very fat-tailed, this prior gives a lot of mass to a wide range of plausible effect sizes while at the same time not reducing the ability to obtain evidence for smaller effects by much (Wagenmakers et al., 2020).

We found no support for the main effect of Identifiability (H1a), Explicit Learning (H3a), or their interaction (H2a), with similar hypothetical donation amounts in the identifiable victim and statistical victim conditions. We therefore concluded failure to replicate the identifiable victim effect (H1a), and failure to replicate that explicitly learning about the effect impacted the effect itself (H2a).

H1b, H2b and H3b: Identifiability and Explicit Learning Main Effects and Interaction (with Joint Condition) [Extension]

We ran an additional more complex version of the analysis above, which included a comparison to the joint condition (which was added in the target article's Study 3) and was only possible because of our unified design combining replications of the target article's Studies 1 and 3. We conducted a 2 (Explicit Learning) × 3 (Identifiability) two-way ANOVA to examine if the provision of additional quantitative information together with an identified victim would debias the identifiable victim effect.

We found no support for the main effect of Explicit Learning and no interaction effect of Identifiability and Explicit Learning, that explicitly learning about the identifiable victim effect reduces people's (hypothetical) donations (H3b). We found some support for main effect of Identifiability (H1b), $F(2, 998) = 3.91$, $p = .02$, $\eta_p^2 = .008$. 95% CI [.000, .021], though Bayesian analysis indicates weak support for the *null* ($BF_{01} = 1.77$). The different conclusions from the two analyses can be explained by the large sample that increases the likelihood of significant *p*-values, even when

4 We had preregistered to check normality and kurtosis of dependent variables. However, we realized that given our large sample size due to central limit theorem the sample means would still be normally distributed even if the data is not and therefore did not conduct these tests.

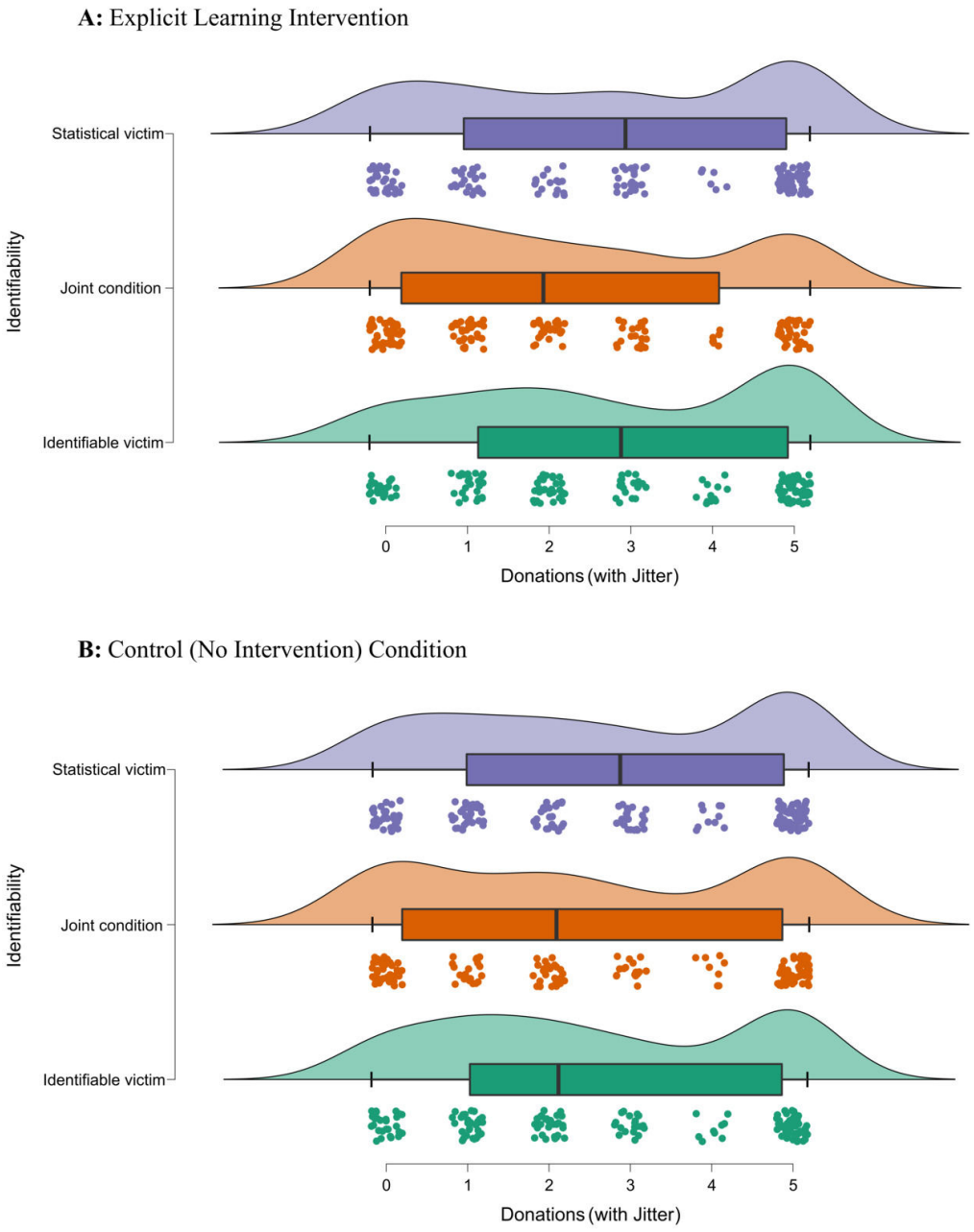


Figure 2. Hypothetical donations: Interaction of Identifiability and Explicit Learning

Note. Created in JASP (2023) version 0.16.

the evidence is low from a Bayesian perspective (Maier & Lakens, 2022).

To better understand the Identifiability main effect, we also examined the post-hoc comparisons comparing the different Identifiability conditions with Bonferroni correction. We found no support for differences between statistical and identifiable victim conditions, $t(998) = 0.097$, $p = 1.00$, $BF_{01} = 11.57$, $d = 0.01$, 95% CI [-0.16, 0.14], and near threshold for the comparison between identifiable and joint, $t(998) = 2.37$, $p = .053$, $d = 0.18$ [0.03, 0.34] with donations slightly lower in the joint condition. We found support for differences between the statistical and the joint condition $t(998) = 2.46$, $p = .041$, $d = 0.19$ [0.04, 0.34]. Given the weak near threshold unexpected effect, we caution against over-interpretation of the Identifiability main effect or the contrasts.

H4: Identifiability with Implicit Learning (Joint Condition) Main Effect (without Explicit Learning) [Replication]

We conducted the analyses mirroring the analyses of Study 3 in Small et al (2007), without including the explicit learning intervention conditions (i.e., H4: Identifiable-Control > Joint-Control ~= Joint-Control). Although this was conducted by the target, it was not included in the pre-

Table 8. Hypothetical Donations: Statistical Tests for Identifiability and Explicit Learning

	F	*p*	BF_{01}	η_p^2	95% CI
H1: Identifiability					
Without joint condition [S1] H1a: Identifiable (Explicit & Control) vs. Statistical (Explicit & Control)					
Target article	6.75	< .05	N/A	.06	[.00, .15]
Replication	0.01	.923	11.57	.00	[.00, .003]
With joint condition [E] H1b: Identifiable (Explicit & Control) vs. Statistical (Explicit & Control) vs. Joint (Explicit & Control)					
Replication	3.91	.020	1.77	.01	[.00, .021]
H2: Interaction: Identifiability and Explicit Learning					
Without joint condition [S1] H2a: (Identifiable-Explicit vs. Statistical-Explicit vs. Identifiable-Control vs. Statistical-Control)					
Target article	5.32	< .05	N/A	.04	[.00, .14]
Replication	0.654	.419	6.30	.001	[.000, .011]
With joint condition [E] H2b: (Identifiable-Explicit vs. Statistical-Explicit vs. Joint-Explicit vs. Identifiable-Control vs. Statistical-Control vs. Joint-Control)					
Replication	1.48	.228	12.74	.003	[.000, .012]
H3: Explicit learning intervention					
Without joint condition [S1] H3a: Explicit (Identifiable & Statistical) vs. Control (Identifiable & Statistical)					
Target article	4.15	< .05	N/A	.03	[.00, .12]
Replication	0.89	.346	7.51	.00	[.000, .012]
With joint condition [E] H3b: Explicit (Identifiable-, Statistical, & Joint) vs. Control (Identifiable, Statistical, & Joint)					
Replication	0.005	.943	14.15	.00	[.000, .002]
H4: Implicit learning and Identifiability					
Without explicit learning [S3] * H4: Identifiable (Control) > Statistical (Control) ~= Joint (Control)					
Target article	5.67	< .01	N/A	.07	[.01, .15]
Replication	0.61	.541	25.03	.00	[.00, .015]

Note. ANOVA tests. $N = 1004$. CI = confidence interval. N/A = could not be recalculated. BF_{01} denotes the Bayes factor in favor of the null. Bayes factors based on Cauchy prior with $r_{scale} = 0.707$. η_p^2 for original study recalculated based on F-statistics and degrees of freedom.
[S1] mirrors the target article's Study 1. [S3] mirrors the target article's Study 3. [E] indicates an extension. * indicates analysis was not pre-registered.

registration, which was focused on the unified design and included the explicit conditions (see below). We therefore labeled this analysis exploratory. We found no support for an implicit learning effect, $F(2, 499) = 0.61$, $p = .541$, $\eta_p^2 = .002$, 95% CI [.00, .02], and with strong evidence against the effect in a complementary Bayesian analysis ($BF_{01} = 25.03$). Therefore, we did not conduct any follow-up tests comparing differences between specific cells.

Feelings

The Cronbach's alpha for the feelings variables was 0.90. We therefore followed the methodology by Small et al (2007) and aggregated the five feelings into a single measure of aggregated feelings, combining: 1) feeling upset, 2) feeling sympathetic towards the victim(s), 3) feeling touched by the situation, 4) feeling morally responsible, and 5) feeling that it is appropriate to donate to the cause. We summarized the results of the hypotheses tested on aggregated feelings in Table 9.

H5a/b/c: Identifiability and Explicit Learning Main Effects and Interaction on Aggregated Feelings (without Joint Condition) [Replication]

In Small et al. (2007), aggregated feelings were measured and analyzed in Study 1, and the joint condition was introduced in Study 3. We therefore first conducted a matched analysis to their Study 1 without the joint condition. We note that our pre-registration originally focused on the analyses that included the joint condition, yet deviated from the target's Study 1, which is reported in the following section.

We found no support for and with Bayesian analyses evidence against the main effect of Identifiability (H5a: $F(1, 666) = 0.09$, $p = .764$, $BF_{01} = 11.08$, $\eta_p^2 = 0.00$, 95% CI [0.00, 0.01]), main effect of Explicit Learning and (H5b: $F(1, 666) = 0.01$, $p = .940$, $BF_{01} = 11.56$, $\eta_p^2 = 0.00$, 95% CI [0.00, 0.002]), and their interaction (H5c: $F(1, 666) = 0.04$, $p = .842$, $BF_{01} = 8.13$, $\eta_p^2 = 0.00$, 95% CI [0.00, 0.01]).

Table 9. Aggregated Feelings: Statistical Tests for Identifiability and Explicit Learning

	df	*F*	*p*	BF_{01}	η_p^2	95% CI
Identifiability main effect, without joint [S1] * H5a: Identifiable (Explicit & Control) > Statistical (Explicit & Control)						
Target article	1, 114	1.80	.18	N/A	.02	[.00, .09]
Replication	1, 114	0.09	.764	11.08	.00	[.00, .01]
Explicit learning intervention main effect, without joint [S1] * H5b: Explicit (Identifiable, Statistical) vs. Control (Identifiable, Statistical)						
Target article	1, 114	0.24	.63	N/A	.00	[.00, .05]
Replication	1, 114	0.01	.940	11.56	.00	[.00, .002]
Interaction effect, without joint [S1] * H5c: (Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control)						
Target article	1, 114	2.00	.16	N/A	.02	[.00, .09]
Replication	1, 114	0.04	.842	8.13	.00	[.00, .01]
Interaction effect, with joint [E] H5d: (Identifiable-Explicit minus Identifiable-Control > Statistical-Explicit minus Statistical-Control ~= Joint-Explicit minus Joint-Control)						
Replication	2, 998	0.792	.453	44.85	.002	[.00, .009]

Note. ANOVA test. $N = 1004$. CI = confidence interval. Aggregated feelings refer to averaging the feelings of being upset, being sympathetic, being touched, moral responsibility, and donation appropriateness into a single composite. The Cronbach's alpha for the five feelings measures was 0.90. BF_{01} denotes the Bayes factor in favor of the null. Bayes factors based on Cauchy prior with $r_{scale} = 0.707$. [S1] mirrors the target article's Study 1. [E] indicates an extension. * indicates analysis was not pre-registered.

H5d: Identifiability and Explicit Learning Main Effects and Interaction on Aggregated Feelings (with Joint Condition) [Extension]

We also ran a pre-registered extension analysis with the joint condition that went beyond the target article's Study 1. We conducted a 2 (Explicit Learning) × 3 (Identifiability) two-way ANOVA on aggregated feelings (mean across all feelings measures apart from 'perceived impact'). Results were similar to those without the joint condition, with no support for the main effects of Identifiability, the main effect of Explicit Learning, or the interaction on aggregated feelings. The Bayesian analysis further suggests evidence against an effect. Thus, we concluded that H5a/b/c and H5d were not supported.

Exploratory: Identifiability and Explicit Learning interaction on singular feelings

We also ran a 2 (Explicit Learning) × 3 (Identifiability) two-way ANOVA to determine if separate affective reactions were affected by the Identifiability and Explicit Learning interaction. We summarized our findings in a table in the supplementary materials in the section 'Descriptive Statistics and Tests for Disaggregated Feelings'.

The target article reported no support for any effects regarding the feeling variables. In this replication, we found some support for a main effect of Identifiability on moral responsibility, $F(2, 998) = 3.82$, $p = .022$, $\eta_p^2 = .008$ [.000, .021] and appropriateness to donate, $F(2, 998) = 3.71$, $p = .025$, $\eta_p^2 = .007$ [.000, .020]. However, a Bayesian analysis suggests weak evidence *against* an effect on moral responsibility ($BF_{01} = 2.05$) and appropriateness of donation ($BF_{01} = 2.24$). We therefore conclude that there is not enough evidence to claim an effect on disaggregated feelings (especially given the large sample size and number of statistical tests).

Associations between Aggregated Feelings and Hypothetical Donations

We examined the associations between five aggregated feelings (as above) and hypothetical donations, summarized in Table 10. The target article found that aggregated feelings predicted donations more strongly in the identifiable victim/no intervention condition than in the other conditions. In our data, there were strong positive relationships between aggregated feelings and hypothetical donations across all six conditions, with the effects in all conditions overlapping with the confidence intervals of all the effects in the other conditions, showing no indication for differences. We concluded these results as inconsistent with the target article's findings.

Perceived impact (Extension)

H6: Effect of Identifiability and Explicit Learning on Perceived Impact of Donation

We ran a 2 (Explicit Learning) x 3 (Identifiability) two-way ANOVA to determine how the perceived impact of the donation differed depending on these two factors. We find evidence for an effect of Identifiability on perceived impact, $F(2, 998) = 10.5$, $p = .00003$, $\eta_p^2 = .021$ [.006, .040], $BF_{01} = 0.003$. However, we do not find evidence for an effect of Explicit Learning on perceived impact, $F(1, 998) = 0.11$, $p = 0.74$, $\eta_p^2 = .000$ [.000, .004], $BF_{01} = 13.69$, or for an interaction between Explicit Learning and Identifiability, $F(2, 998) = 1.48$, $p = 0.229$, $\eta_p^2 = .002$ [.000, .012], $BF_{01} = 11.05$.

Table 10. Correlations between Aggregated Feelings and Hypothetical Donations across Conditions

	Target article			Replication		
Conditions	*r*	*p*	*n*	*r*	95% CI	*p*
Identifiable/ Explicit learning	.34	N/A	170	.63	[0.53, 0.72]	< .001
Identifiable/ No explicit learning	.55	< .01	165	.58	[0.47, 0.67]	< .001
Statistical/ Explicit learning	.33	N/A	159	.64	[0.54, 0.73]	< .001
Statistical/ No explicit learning	.39	N/A	176	.56	[0.45, 0.65]	< .001
Joint/ Explicit learning	N/A	N/A	173	.63	[0.53, 0.71]	< .001
Joint/ No explicit learning	N/A	N/A	161	.59	[0.58, 0.80]	< .001

Note. CI = confidence interval. N/A = unreported in the original studies. Aggregated feelings refer to averaging the feelings of being upset, being sympathetic, being touched, moral responsibility, and donation appropriateness into a single composite. The Cronbach's alpha for the five feelings measures was 0.90.

As we only found evidence for an effect of Identifiability, we follow up with post-hoc tests of this factor. We found support for higher perceived impact in the identifiable victim condition (M = 3.39, SD = 1.24) compared to both the statistical victim (M = 2.96, SD = 1.35), $t(668) = 4.35$, $p < .001$, $d = 0.34$, 95% CI [0.18, 0.49], and joint condition (M = 3.02, SD = 1.39), $t(667) = 3.67$, $p < .001$, $d = 0.28$, 95% CI [0.13, 0.44]. Both of these results also held up in the Bayesian analysis (BF_{01} = 0.001 and 0.02).

Further, the perceived impact of donation was correlated to hypothetical donations, $r(1002) = 0.54$, $p < .001$. The correlation is comparable for all cells of our design ranging from 0.48 in the identifiable victim/no intervention condition to 0.60 in the statistical victim/intervention condition (see supplementary materials for the correlation in each cell of our design).

Discussion

We conducted a replication and extension of Small et al.'s (2007) Studies 1 and 3 to examine the discrepancy in human hypothetical prosocial donations towards a single identifiable victim compared to a group of anonymous statistical victims. We found no support for the identifiable victim effect in hypothetical donation tasks and Bayesian analyses indicated evidence in support of no effect. We found further support for this null effect in our reanalysis of a large meta-analysis on the effect conducted by Lee and Freely (2016) using advanced publication bias adjustment methods. We also failed to demonstrate that either explicitly learning about the effect explicitly (through reading prior research) or implicitly (being given statistical victim information along with the personalized victim) weakens the hypothetical donations gap. Thus, we conclude that we failed to find support for the target article's findings regarding the identifiable victim effect and the interventions that weakened the effect in hypothetical donations. We provided a comparison of the results between the original study and replication in Table 11.

In our extension adding a measure of perceived impact, we found support for perceived impact of the hypothetical donations as higher for an identifiable victim compared to statistical victims, and with support for an association between perceived impact and hypothetical donations, though it somehow failed to translate to an effect on hypothetical donations. Further research is needed to try and understand the links between perceived impact, hypothetical donations, intent to donate, and actual donations.

We caution that our results should not be considered a 'final word' on this effect but rather a motivation for future replication efforts in the form of high-powered registered reports examining hypothetical donations, donation intent, real money donations, and associated perceptions such as perceived impact. In addition, we see many promising theoretical directions for further work in this area and possibilities for rethinking and reframing the original theory.

Identifiable Victim Effect or Scope Insensitivity?

Majumder et al. (2023) recently reported a failed replication of Kogut and Ritov (2005) and suggested that the identifiable victim effect may be reframed, that instead of larger donations towards an identifiable victim, the effect might be viewed as similar donations towards an identifiable victim as a group of unidentified or statistical victims with no donation adjustment per group size. This cognitive phenomenon is usually discussed under the term 'scope insensitivity', and describes that people do not value a good (here helping children in need) in proportion to its scope or size (Baron & Greene, 1996; Desvousges et al., 1993; Kahneman & Knetsch, 1992). Scope insensitivity has also been shown to be a factor in charitable giving (Hsee et al., 2013; Maier, Caviola, et al., 2022; Västfjäll & Slovic, 2020) and has been discussed as a reason for neglecting to help save human lives, for example, in the context of genocides (Cameron & Payne, 2011; Dickert et al., 2012, 2015; Slovic & Västfjäll, 2010). We see much need for research that would help clarify the different aspects of the phenomenon, to disentangle Identifiability (whether targets are identified or not), from singularity (one versus group), from group size (in scope insensitivity), and to then revisit the classics and examine each of these factors, separately and jointly. Across several replications, we struggled to find support for seminal articles in this domain (most recently, in Mayiwar et al., 2023), and it would seem that these challenges are also shared by the very scholars who initially reported these phenomena (e.g., Moche & Västfjäll, 2021)

Table 11. Replication Results Summary: Comparison between the Target Article and the Replication

H	Study	Independent variable	Dependent variable	η_p^2 effect size and 95% Confidence intervals		Replication summary	BF_{01} Evidence for null
				Target article	Replication		
1a	1	Identifiability main effect, without joint	Donation	.06 [.00, .15]	.00 [.00, .003]	No signal – inconsistent	11.57
2a	1	Identifiability x Explicit Learning interaction effect, without joint	Donation	.04 [.00,.14]	.001 [.00, .011]	No signal – inconsistent	6.30
3a	1	Explicit Learning main effect, without joint	Donation	.03 [.00, .12]	.001 [.00, 0.01]	No signal - inconsistent	7.51
4	3	Implicit learning main effect, without Explicit	Donation	.07 [.01, .15]	.00 [.00, .015]	No signal - inconsistent	25.03
5a	1	Identifiability main effect, without joint	Affective feelings	0.02 [.00, .09]	.00 [.00, .01]	No signal - inconsistent	11.08
5c	1	Identifiability x Explicit Learning interaction effect, without joint	Affective feelings	0.02 [.00, .09]	.00 [.00, 0.006]	No signal - inconsistent	8.13

Note. H = Hypotheses. Replication summary is using the LeBel et al. (2019) criteria.

Evidence for Irrational Decision Making?

It is unclear whether this phenomenon can be considered evidence for irrational decision-making in the context of identifiable victims. On the one hand, as Majumder et al. (2023) argued, the larger group of victims should elicit more empathic concern, distress, and consequently, willingness to contribute. Not observing this pattern violates the principle of proportionality (i.e., larger issues should be tackled with more resources). On the other hand, from a cost-effectiveness perspective, it makes sense to contribute more where the donation is most *effective* rather than where the problem is biggest. According to the theory of impact philanthropy proposed by Duncan (2004), the tendency for people to offer help lies in their perception of the difference they can make with their donations. In our study, we found that participants did not necessarily perceive a hypothetical donation to the larger group as more impactful, but rather that they may in fact consider donations to the identifiable victim more impactful, in line with Duncan (2004). People might also perceive donating to the identified victim as more impactful due to proportion dominance. In other words, they may donate less to statistical victims, given that they perceive a lower impact of their contribution when they can only help a smaller proportion of affected individuals (e.g., Erlandsson et al., 2014).

Therefore, effectiveness-based reasoning would imply the opposite compared to the principle of proportionality – donating more to the identified victim. A potential explanation of the null effects in our study would be that participants apply both reasoning based on proportionality and based on effectiveness, and the two cancel each other out, resulting in an overall null effect. Future research may measure participants' effectiveness focus and tendency to allocate resources based on proportionality to directly investigate how these two factors affect donations to the identifiable victim.

Limitations and More Future Directions

A core limitation that may explain the discrepancy between the results of the original studies and our replication is our adjustment from real to hypothetical donations. In Small et al. (2007), participants received money as a reward after filling in an unrelated survey about the use of various technology products. Participants then received a blank envelope and a charity request letter to decide how much they would be willing to donate. Answering the unrelated technology survey allowed participants to assess how much effort they invested to earn money, making it easier to grasp the subjective value of the money than in our study. Second, given this cover story, participants may not have realized that the experimenters were investigating their donation behavior. Third, participants might donate differently with real in comparison to imaginary money, as they would, for instance, likely deliberate more when making choices involving real donations.

In our replication, we asked the participants to imagine they had just earned $5 and how much of this they would like to give to the corresponding victims. Generosity reflected in the hypothetical donation is usually higher than that expected in the original studies (Bekkers, 2006). Though a direct comparison between the two studies is problematic given the passing of time and the very different measures, looking at the raw numbers in our replication people indicated higher hypothetical donations (Table 5 in the 'Results' section), compared to the real donations reported in the target article. However, we note that our conclusions do not depend on average donations but on the differences between conditions. We are not aware of

any evidence that would suggest that these effects stand a better chance of working in real life setting than they do in hypothetical scenarios. Nevertheless, a replication in a field setting or an experiment with real donations would be valuable in the future, though we recommend adjusting expectations and taking into account that observed effects might be much weaker than initially thought.

Second, we made additional adjustments and also added forced comprehension checks, to ensure that participants read and understood the hypothetical donation situation and choice. It is possible that this may have somehow impacted participants' responses since they might disrupt feelings of empathy. In addition, participants may believe that the information about the identified victim effect was supplied to them in order to answer the comprehension checks rather than in order to use it in the subsequent donation task. However, we note that if the effect was indeed affected by such factors, it may indicate that the initial demonstrations were atleast partially motivated by socially desirability responding (McKenzie et al., 2018), and/or that the effect is more contextual, weaker, and less robust than initially thought.

Third, our study was conducted online rather than in person (as in Small et al., 2007). On the one hand, this difference may also be considered a strength, as the online data collection allowed us to collect a larger and broader sample than would have been possible in a lab study. On the other hand, the increased anonymity in online settings could reduce participants' willingness to donate, even though it is less clear how this would affect the differences between conditions. This research was also conducted during the Covid-19 pandemic, which might have affected participants' financial status and their psyche more broadly. These two factors might have resulted in our participants having little money for donations or being pre-occupied with financial and existential concerns. Hypothetical donations, therefore, may have been limited by resource constraints, or their 'mental account' of how much they are willing to contribute to donation tasks (Sussman et al., 2015; Thaler, 1985, 1999).

Conclusion

We conducted a replication and extension of Small et al. (2007) with modified setting and using hypothetical donations. Contrary to the target article's findings, we did not find support for the identifiable victim effect and did not find support for explicit and implicit interventions as weakening the effect. We emphasize that our paper should not be considered conclusive evidence against the identifiable victim effect, given the differences in the experimental setup. Instead, we believe that the failure to find the effect on hypothetical donations in combination with the publication bias-adjusted meta-analysis constitutes a cautionary note. We, therefore, conclude that our paper shows a pressing need for more replications with real donations in the form of registered replication reports (Chambers, 2013), ideally conducted as adversarial collaborations between proponents and critics of the identifiable victim effect.

Competing Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

Author Contributions

Maximilian Maier built on the thesis work by Yik Chun, verified all analyses, added additional analyses (Bayesian), the RoBMA reanalysis of Lee and Freely (2016), new visualizations, and wrote an initial journal submission manuscript.

Yik Chun Wong conducted the replication as part of her dissertation.

Gilad was the advisor for the dissertation. Gilad supervised each step in the project, conducted the preregistrations, and ran data collection.

Maximilian and Gilad finalized the journal submissions, revised and responded to peer review.

Acknowledgements

We thank František Bartoš for conducting an independent verification report of the RoBMA analysis.

We thank Vanessa Cheung for helpful feedback on a previous draft of this manuscript.

Submitted: December 24, 2022 PST, Accepted: October 15, 2023 PST

Contributor Roles Taxonomy

Role	Maximilian Maier	Yik Chun Wong	Gilad Feldman
Conceptualization			X
Pre-registration		X	X
Data curation			X
Formal analysis	X	X	
Funding acquisition			X
Investigation	X	X	
Pre-registration peer review / verification			X
Data analysis peer review/verification	X		
Methodology	X	X	X
Project administration			X
Resources			X
Software	X	X	
Supervision			X
Validation	X		
Visualization	X	X	
Writing-original draft	X	X	
Writing-review and editing	X		X

References

Alinaghi, N., & Reed, W. R. (2018). Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? *Research Synthesis Methods*, *9*(2), 285–311. https://doi.org/10.1002/jrsm.1298

Banerjee, A., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, *14*(3), 301–309. https://doi.org/10.1023/a:1007796310463

Baron, J., & Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*, *2*(2), 107–125. https://doi.org/10.1037/1076-898x.2.2.107

Bartoš, F., & Maier, M. (2020). *RoBMA: An R package for robust Bayesian meta-analyses*. R package version 2.1.1. https://CRAN.R-project.org/package=RoBMA

Bartoš, F., Maier, M., Quintana, D. S., & Wagenmakers, E.-J. (2022). Adjusting for publication bias in JASP and r — Selection models, PET-PEESE, and robust Bayesian meta-analysis. *Advances in Methods and Practices in Psychological Sciences*. In press. https://doi.org/10.31234/osf.io/75bqn

Bartoš, F., Maier, M., Shanks, D., Stanley, T. D., Sladekova, M., & Wagenmakers, E.-J. (2022). *Meta-Analyses in Psychology Often Overestimate Evidence for and Size of Effects*. https://doi.org/10.31234/osf.io/tkmpc

Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2022). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*. Advance online publication. https://doi.org/10.1002/jrsm.1594

Bekkers, R. (2006). *Words and Deeds of Generosity: Are Decisions About Real and Hypothetical Money Really Different*. Working paper, Department of Sociology, Utrecht University. https://renebekkers.files.wordpress.com/2015/12/15_12_01_words_and_deeds.pdf

Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973. https://doi.org/10.1177/0899764010380927

Bergh, R., & Reinstein, D. (2021). Empathic and numerate giving: The joint effects of victim images and charity evaluations. *Social Psychological and Personality Science*, *12*(3), 407–416. https://doi.org/10.1177/1948550619893968

Bom, P. R., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*(4), 497–514. https://doi.org/10.1002/jrsm.1352

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, *100*(1), 7–42. https://doi.org/10.1023/a:1007850605129

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Caviola, L., Schubert, S., & Nemirow, J. (2020). The many obstacles to effective giving. *Judgment & Decision Making*, *15*(2), 159–172. https://doi.org/10.1017/s1930297500007312

Chambers, C. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. (2017). *The seven deadly sins of psychology*. Princeton University Press.

Charris, R. A. (2018). *A Systematic Replication of the identifiable victim effect (Small, Loewenstein & Slovic 2007)* [Unpublished doctoral dissertation, University of Los Andes (Colombia)]. https://repositorio.uniandes.edu.co/bitstream/handle/1992/34740/u808458.pdf?sequence=1

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Hudson, S. P., Wilson, K. N., & Boyle, K. J. (1993). Measuring natural resource damages with contingent valuation: Tests of validity and reliability. In B. H. Baltagi & F. Moscore (Eds.), *Contributions to Economic Analysis* (pp. 91–164). Elsevier. https://doi.org/10.1016/b978-0-444-81469-2.50009-2

Dickert, S., Västfjäll, D., Kleber, J., & Slovic, P. (2012). Valuations of human lives: Normative expectations and psychological mechanisms of (ir)rationality. *Synthese*, *189*(S1), 95–105. https://doi.org/10.1007/s11229-012-0137-4

Dickert, S., Västfjäll, D., Kleber, J., & Slovic, P. (2015). Scope insensitivity: The limits of intuitive valuation of human lives in public policy. *Journal of Applied Research in Memory and Cognition*, *4*(3), 248–255. https://doi.org/10.1016/j.jarmac.2014.09.002

Duncan, B. (2004). A theory of impact philanthropy. *Journal of Public Economics*, *88*(9–10), 2159–2180. https://doi.org/10.1016/s0047-2727(03)00037-9

Erlandsson, A., Björklund, F., & Bäckström, M. (2014). Perceived utility (not sympathy) mediates the proportion dominance effect in helping decisions. *Journal of Behavioral Decision Making*, *27*(1), 37–47. https://doi.org/10.1002/bdm.1789

Erlandsson, A., Björklund, F., & Bäckström, M. (2015). Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively. *Organizational Behavior and Human Decision Processes*, *127*, 1–14. https://doi.org/10.1016/j.obhdp.2014.11.003

Friedrich, J., & McGuire, A. (2010). Individual differences in reasoning style as a moderator of the identifiable victim effect. *Social Influence*, *5*(3), 182–201. https://doi.org/10.1080/15534511003707352

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance Decision Making: A Field Study of Prosocial Lending. *Journal of Marketing Research*, *48*(SPL), S130–S137. https://doi.org/10.1509/jmkr.48.spl.s130

Hart, P. S., Lane, D., & Chinn, S. (2018). The elusive power of the individual victim: Failure to find a difference in the effectiveness of charitable appeals focused on one compared to many victims. *Plos One*, *13*(7), e0199535. https://doi.org/10.1371/journal.pone.0199535

Hong, S., & Reed, W. R. (2021). Using Monte Carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, *12*(2), 192–215. https://doi.org/10.1002/jrsm.1467

Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological Science*, *24*(9), 1801–1808. https://doi.org/10.1177/0956797613482947

Isager, P. M., van Aert, R. C. M., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., Krueger, J. I., Perugini, M., Ropovik, I., van't Veer, A. E., Vranka, M., & Lakens, D. (2021). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000438

JASP Team. (2023). *JASP* (Version 0.16).

Jeffreys, H. (1939). *Theory of probability* (1st ed.). Oxford University Press.

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, *14*(3), 235–257. https://doi.org/10.1023/a:1007740225484

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, *22*(1), 57–70. https://doi.org/10.1016/0095-0696(92)90019-s

Kogut, T., & Ritov, I. (2005). The "identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/10.1002/bdm.492

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, *333*(7568), 597–600. https://doi.org/10.1136/bmj.333.7568.597

Lebel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. https://doi.org/10.1177/2515245918787489

Lebel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, *3*, 1–9. https://doi.org/10.15626/mp.2018.843

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lee, S., & Feeley, T. H. (2016). The identifiable victim effect: a meta-analytic review. *Social Influence*, *11*(3), 199–215. https://doi.org/10.1080/15534510.2016.1216891

Lee, S., & Feeley, T. H. (2018). The identifiable victim effect: Using an experimental-causal-chain design to test for mediation. *Current Psychology*, *37*(4), 875–885. https://doi.org/10.1007/s12144-017-9570-3

Lesner, T. H., & Rasmussen, O. D. (2014). The identifiable victim effect in charitable giving: evidence from a natural field experiment. *Applied Economics*, *46*(36), 4409–4430. https://doi.org/10.1080/00036846.2014.962226

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Loewenstein, G. F., Small, D. A., & Strnad, J. F. (2006). Statistical, identifiable, and iconic victims. *Behavioral Public Finance*, *32–46*, 32–46. https://doi.org/10.2139/ssrn.678281

Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2022). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000405

Maier, M., Caviola, L., Schubert, S., & Harris, A. J. L. (2022). *Investigating (Sequential) Unit Asking: An Unsuccessful Quest for Scope Sensitivity in Willingness to Donate Judgements*. https://doi.org/10.31234/osf.io/ps34b

Maier, M., & Lakens, D. (2022). Justify Your Alpha: A Primer on Two Practical Approaches. *Advances in Methods and Practices in Psychological Science*. In press. https://doi.org/10.31234/osf.io/ts4r6

Maier, M., VanderWeele, T. J., & Mathur, M. B. (2022). Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. *Campbell Systematic Reviews*, *18*(3), e1256. https://doi.org/10.1002/cl2.1256

Majumder, R., Tai, Y. L. C., Ziano, I., & Feldman, G. (2023). *Revisiting the impact of singularity on the Identified Victim Effect: An unsuccessful replication and extension of Kogut and Ritov (2005a) Study 2*. https://osf.io/9qcpj/

Mayiwar, L., Rudko, I., Jeong, Y., Rekdal, R., Do, T. V., Yu, S., Dorotic, M., & Feldman, G. (2023). Replication and extensions of Experiments 1a and 3 in Vastfall et al. (2014). *Open Science Framework*. https://doi.org/10.17605/OSF.IO/A2TGB

McKenzie, C. R., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics*, *5*(3–4), 337–370. https://doi.org/10.1561/105.00000091

Moche, H. (2022). *Unequal Valuations of Lives and What to Do About It: The Role of Identifiability, Numbers, and Age in Charitable Giving* [Doctoral dissertation, Linköping University Electronic Press]. https://doi.org/10.3384/9789179295226

Moche, H., & Västfjäll, D. (2021). Helping the child or the adult? Systematically testing the identifiable victim effect for child and adult victims. *Social Influence*, *16*(1), 78–92. https://doi.org/10.1080/15534510.2021.1995482

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research: A comparative evaluation of six statistical methods. *Zeitschrift für Psychologie*, *227*(4), 261–279. https://doi.org/10.1027/2151-2604/a000386

Ritov, I., & Kogut, T. (2011). Ally or adversary: The effect of identifiability in inter-group conflict situations. *Organizational Behavior and Human Decision Processes*, *116*(1), 96–103. https://doi.org/10.1016/j.obhdp.2011.05.005

Singer, P. (2019). *The life you can save: How to do your part to end world poverty*. The Life You Can Save. org.

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, *2*(2), 79–95. https://doi.org/10.1017/s1930297500000061

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Slovic, P., & Västfjäll, D. (2010). Affect, moral intuition, and risk. *Psychological Inquiry*, *21*(4), 387–398. https://doi.org/10.1080/1047840x.2010.521119

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, *26*(1), 5–16. https://doi.org/10.1023/a:1022299422219

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153. https://doi.org/10.1016/j.obhdp.2006.01.005

Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of Marketing Research*, *46*(6), 777–787. https://doi.org/10.1509/jmkr.46.6.777

Smith, R. W., Faro, D., & Burson, K. A. (2013). More for the many: The influence of entitativity on charitable giving. *Journal of Consumer Research*, *39*(5), 961–976. https://doi.org/10.1086/666470

Sudhir, K., Roy, S., & Cherian, M. (2016). Do sympathy biases induce charitable giving? The effects of advertising content. *Marketing Science*, *35*(6), 849–869. https://doi.org/10.1287/mksc.2016.0989

Sussman, A. B., Sharma, E., & Alter, A. L. (2015). Framing charitable donations as exceptional expenses increases giving. *Journal of Experimental Psychology: Applied*, *21*(2), 130–139. https://doi.org/10.1037/xap0000047

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214. https://doi.org/10.1287/mksc.4.3.199

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*(3), 183–206. https://doi.org/10.1002/(sici)1099-0771(199909)12:3

Västfjäll, D., & Slovic, P. (2020). A psychological perspective on charitable giving and monetary donations: The role of affect. In T. Zaleskiewicz & J. Traczyk (Eds.), *Psychological perspectives on financial decision making* (pp. 331–345). Springer. https://doi.org/10.1007/978-3-030-45500-2_14

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/bf02294384

Wagenmakers, E. J., Ravenzwaaji, D., & Ron, J. (2020, May 14). *Concerns about the default Cauchy are often exaggerated: A demonstration with JASP 0.12*. Bayesian Spectacles. https://www.bayesianspectacles.org/concerns-about-the-default-cauchy-are-often-exaggerated-a-demonstration-with-jasp-0-12/

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*(1), 92–107. https://doi.org/10.1006/jmps.1999.1278

Supplementary Materials

Ported Response Letter

Download: https://collabra.scholasticahq.com/article/90203-revisiting-and-rethinking-the-identifiable-victim-effect-replication-and-extension-of-small-loewenstein-and-slovic-2007/attachment/187057.pdf?auth_token=iSK4UbcI5AmTlauFRvTu

Ported Decision Letter

Download: https://collabra.scholasticahq.com/article/90203-revisiting-and-rethinking-the-identifiable-victim-effect-replication-and-extension-of-small-loewenstein-and-slovic-2007/attachment/187058.pdf?auth_token=iSK4UbcI5AmTlauFRvTu

Peer Review History

Download: https://collabra.scholasticahq.com/article/90203-revisiting-and-rethinking-the-identifiable-victim-effect-replication-and-extension-of-small-loewenstein-and-slovic-2007/attachment/187059.docx?auth_token=iSK4UbcI5AmTlauFRvTu

Supplementary Material

Download: https://collabra.scholasticahq.com/article/90203-revisiting-and-rethinking-the-identifiable-victim-effect-replication-and-extension-of-small-loewenstein-and-slovic-2007/attachment/187060.docx?auth_token=iSK4UbcI5AmTlauFRvTu

Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)

Supplementary

Contents

Open Science Disclosures

Data collection

Data collection was completed before analyzing the data.

Conditions reporting

All collected conditions are reported.

Variables reporting

All variables collected for this study are reported and included in the provided data.

Analysis of the Original Article

Methods of the Original Article

This section summarizes the methodology adopted in Small, Loewenstein, and Slovic (2007)'s Studies 1 and 3

Experimental Design

Study 1:

- 2x2 between-subject design
- Participants were randomly assigned to each condition

Table S1
Study 1: Experimental Design

<table>
<tr><td>IV1: Identifiability [between-subject]

IV2: Intervention [between-subject]</td><td><u>IV1: Identifiable victim condition</u>

Manipulation: Information of an identifiable victim was presented</td><td><u>IV1: Statistical victim condition</u>

Manipulation: Information of statistical victims was presented</td></tr>
<tr><td><u>IV2: Intervention condition</u>
Manipulation: Research on identifiability was presented</td><td colspan="2" rowspan="2">(1 = Not at all, 5 = Extremely)

<u>DV1: Donation</u>
"Indicate the amount of donation from $0 to $5 on the charity request letter and include the letter with the money in the envelope"

<u>DV2: Affective and moral reaction (Feelings)</u>
Feeling of being upset
"How upsetting is this situation to you?"
Feeling of being sympathetic
"How sympathetic did you feel while reading the description of the cause?"
Feeling of being touched
"How touched were you by the situation described?"
Moral responsibility
"How much do you feel it is your moral responsibility to help out with this cause?"
Donation appropriateness
To what extent do you feel that it is appropriate to give money to aid this cause?</td></tr>
<tr><td><u>IV2:No intervention condition</u>
Manipulation: Research on identifiability was not presented</td></tr>
</table>

Note: IV= Independent variable. DV= Dependent variable.

Study 3:

- 1x3 between-subject design

- Participants were randomly assigned to each condition

Table S2
Study 3: Experimental Design

IV1: Identifiability [between-subject]	**IV1: Identifiable victim condition** Manipulation: Information of an identifiable victim was presented	**IV1: Statistical victim condition** Manipulation: Information of statistical victims was presented	**IV1: Joint condition (identifiable victim provided with victim statistics)** Manipulation: Information of statistical and identifiable victims was presented
DV1: Donation "Indicate the amount of donation from $0 to $5 on the charity request letter and include the letter with the money in the envelope"			

Note: IV= Independent variable. DV= Dependent variable.

Independent Variables (IV)

IV1: Identifiability

- Participants were presented with either one of the three victim information indicated below

1. IV1: Identifiable victim condition
    - Manipulation: Information of an identifiable victim was presented

2. IV1: Statistical victim condition
    - Manipulation: Information of statistical victims was presented

3. IV1: Joint condition
    - Manipulation: Information of statistical and identifiable victims was presented

IV2: Intervention

- Participants were assigned to either the intervention or control group. Only participants under manipulation would learn about the identifiable victim effect prior to donations

Dependent Variables (DV)

DV1: Donations

DV2: Affective and moral reactions (feelings)

- Scale: 5-point Likert scale (1 = not at all, 2, 3, 4, 5 = extremely)
- Reliability: $\alpha = .87$
- Feelings variables:

    a. *Feeling of being upset*

    Question: How upsetting is this situation to you?

    b. *Feeling of being sympathetic*

    Question: How sympathetic did you feel while reading the description of the cause?

    c. *Feeling of being touched*

    Question: How touched were you by the situation described?

    d. *Moral responsibility*

    Question: How much do you feel it is your moral responsibility to help out with this cause?

    e. *Donation appropriateness*

    Question: To what extent do you feel that it is appropriate to give money to aid this cause?

Results of the Target Article

Sample Descriptive Statistics

Table S3
Sample Description of the Original Study

Demographics	Small et al. (2007)
Sample size	Study 1: 121 (No exclusion) Study 3: 159 (No exclusion)
Geographic origin	USA
Gender	Unreported
Age	Unreported
Location	A student center at a university in Pennsylvania
Sample type (participants)	Individuals sitting alone in the student center
Sample nature	Field

Statistical Results of the Target Article

Table S4 summarizes the main effects of identifiability and intervention on donations, combining the results of Studies 1 and 3.

Table S4
Effect of Identifiability and Intervention on Donations

Variations	*df1*	*df2*	*F*	*p*	η_p^2
Identifiability					
Study 1: without joint condition with explicit	1	115	6.75	< .05	.06
Study 3: with joint condition, without explicit	2	N/A	5.67	< .01	.07
Explicit Learning	1	115	4.15	< .05	.04
Interaction	1	115	5.32	< .05	.04

Note. ANOVA test. $N = 121$ for identifiability without joint condition. $N = 159$ for identifiability with joint condition. The main effect of intervention and the interaction effect apply to Study 1 but not Study 3. N/A = unreported in the original studies.

Table S5 summarizes the contrast between identifiable/no intervention condition on donations and a combination of other conditions including statistical/no intervention, identifiable/intervention, and statistical/intervention conditions.

Table S5
Mean Contrast of Identifiable/ No Intervention Condition on Donations with a Combination of Other Conditions (Study 1)

Conditions	*M*	*SD*	*df*	*t*	*p*
Identifiable/no intervention	2.83	2.10	117	-4.06	< .001
Combination of other conditions	1.26	1.74			

Note. Independent t-test. $N = 121$.

Table S6 summarizes the ordered probit regression results on the effects of identifiability and intervention on donations. This test was conducted due to the violation of normality with many participants donating $0 to corresponding victims.

Table S6
Effects of Identifiability and Intervention on Donations (Study 1)

Variations	*df*	χ^2	*p*
Identifiability	1	10.06	< .01
Intervention	1	0.01	.92
Interaction	1	4.72	< .03

Note. Ordered probit regression. $N = 121$.

Table S7 summarizes the insignificant effects of identifiability and intervention on aggregated feelings by averaging five segregated feelings into a single composite. Original studies also found insignificant results for the effect of identifiability and intervention on each of these segregated feelings including the feelings of being sympathetic, being upset, being touched, donation appropriateness, and moral responsibility. Results were not reported in the original study.

Table S7
Effect of Identifiability and Intervention on Aggregated Feelings (Study 1)

Variations	*df1*	*df2*	*F*	*p*
Identifiability	1	114	1.80	.18
Intervention	1	114	0.24	.63
Interaction	1	114	2.00	.16

Note. ANOVA test. $N = 121$.

Table S8 summarizes the correlation between aggregated feelings and donations.

Table S8
Study 1: Correlations between Aggregated Feelings and Donations

Conditions	*r*	*p*
Identifiable/ intervention	.34	N/A
Statistical/ intervention	.33	N/A
Statistical/ no intervention	.39	N/A
Identifiable/ no intervention	.55	< .01

Note. Correlation test. $N = 121$. N/A = unreported in the original studies.

Table S9 summarizes the Bonferroni-adjusted pairwise comparison of identifiability on donations in Study 3. Statistical results were not reported in the original study except the *p*-value.

Table S9
Study 3: Effect of Identifiability on Donations with Bonferroni-adjusted Pairwise Comparisons

Identifiability	*p*
Identifiable & Statistical victim	< .01
Statistical victim & Joint condition	1.0
Identifiable victim & Joint condition	< .05

Note. Independent t-test. $N = 159$. *p*-values reported were Bonferroni-adjusted.

Effect Size and Confidence Interval Calculations of the Original Study Effects

Please see the Rmardown or html file in the supplementary materials.

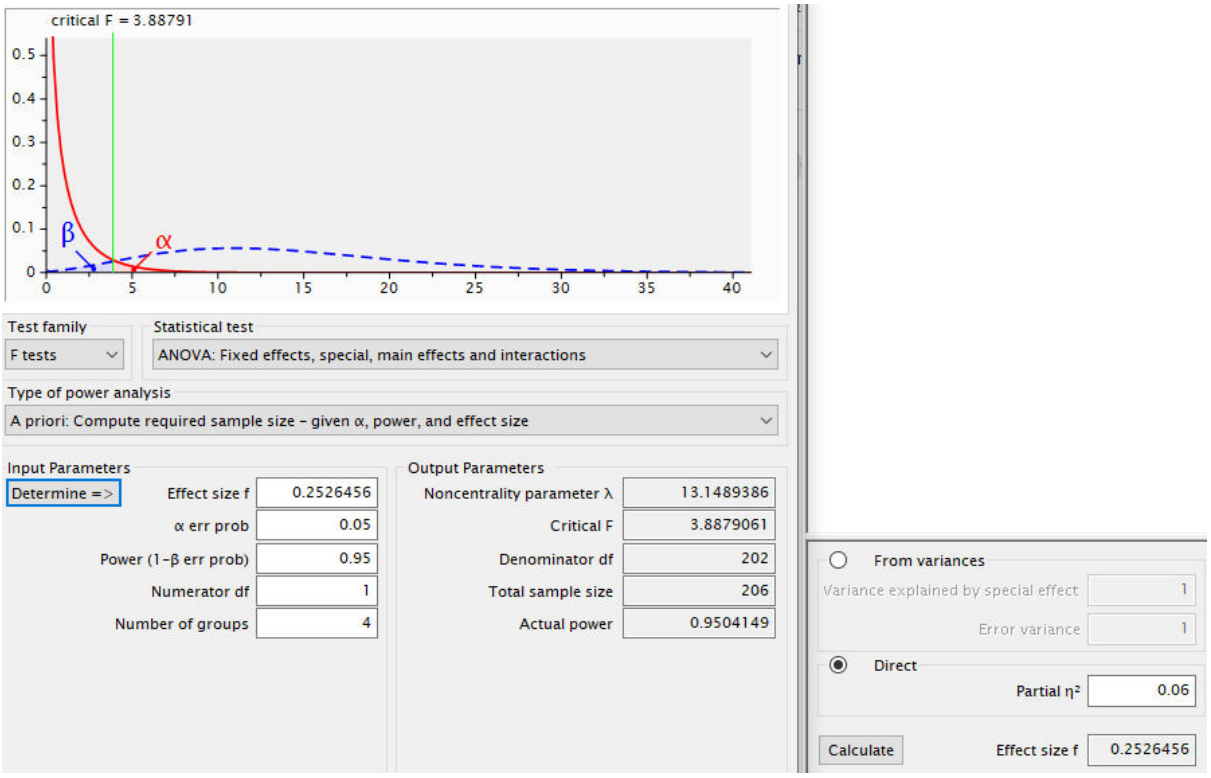
Power Analysis of the Original Study Effect

Power analysis was conducted with G*Power 3.1.9.4 to calculate the sample size for replication. The required sample size for .95 power and .05 alpha is 314.

Main Effect of Identifiability on Donations (Study 1)

Original result: $F(1,115) = 6.75$, $p < .05$, $\eta_p^2 = .06$

- $\eta_p^2 = .06$ obtained from the original study was directly used for calculating the sample size



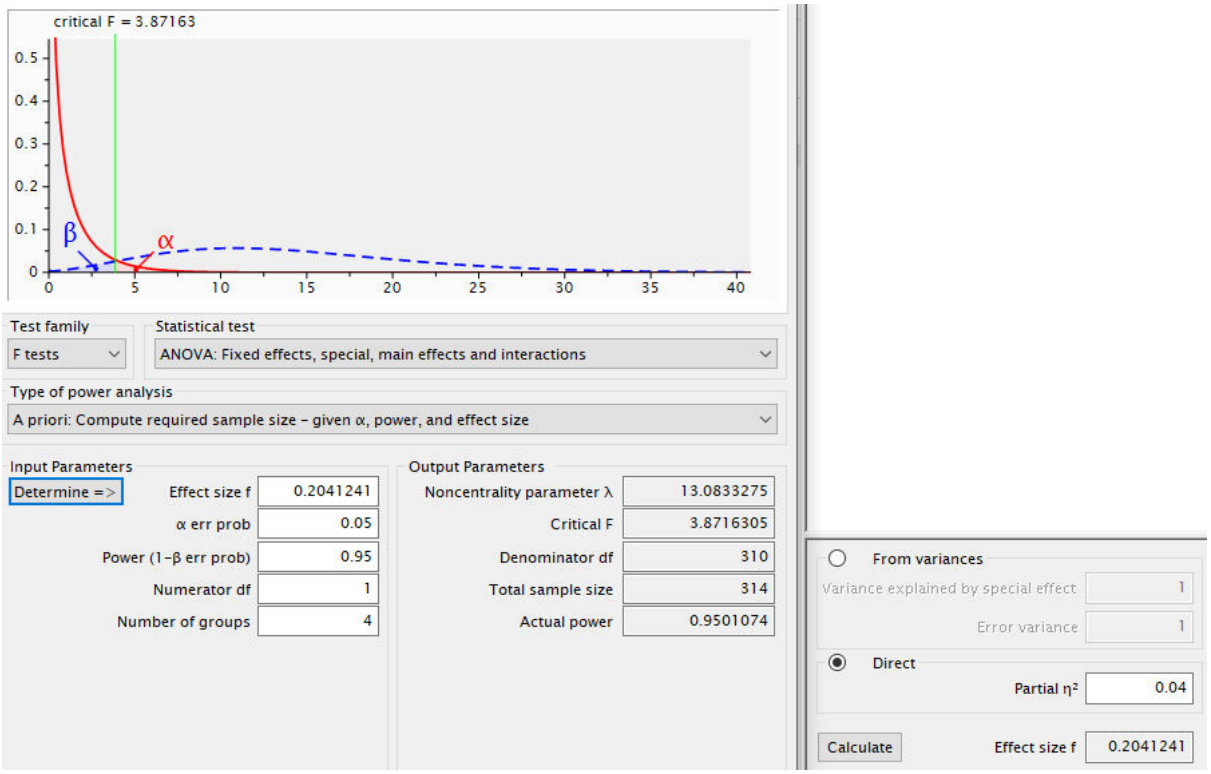
F tests - ANOVA: Fixed effects, special, main effects and interactions
Analysis: A priori: Compute required sample size
Input: Effect size f = 0.2526456
α err prob = 0.05
Power (1-β err prob) = 0.95
Numerator df = 1
Number of groups = 4
Output: Noncentrality parameter λ = 13.1489386
Critical F = 3.8879061
Denominator df = 202

Total sample size = 206
Actual power = 0.9504149

Main Effect of Intervention on donation (Study 1)

Original result: $F(1,115) = 4.15$, $p < .05$, $\eta_p^2 = .04$

- $\eta_p^2 = .04$ obtained from the original study was directly used for calculating the sample size

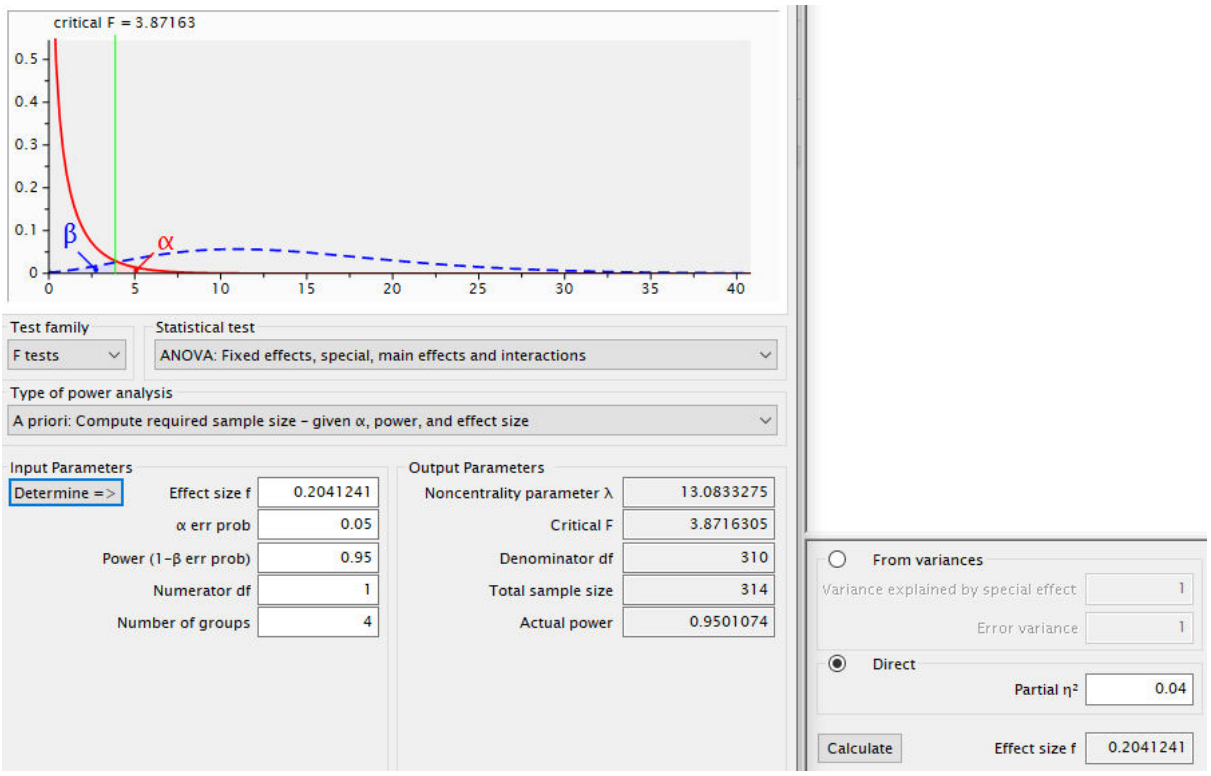


F tests - ANOVA: Fixed effects, special, main effects and interactions
Analysis: A priori: Compute required sample size
Input: Effect size f = 0.2041241
α err prob = 0.05
Power (1-β err prob) = 0.95
Numerator df = 1
Number of groups = 4
Output: Noncentrality parameter λ = 13.0833275
Critical F = 3.8716305
Denominator df = 310
Total sample size = 314
Actual power = 0.9501074

Interaction Effect of Identifiability and Intervention on Donation (Study 1)

Original result: $F(1,115) = 5.32$, $p < .05$, $\eta_p^2 = .04$

- η_p^2 = .04 obtained from the original study was directly used for calculating the sample size

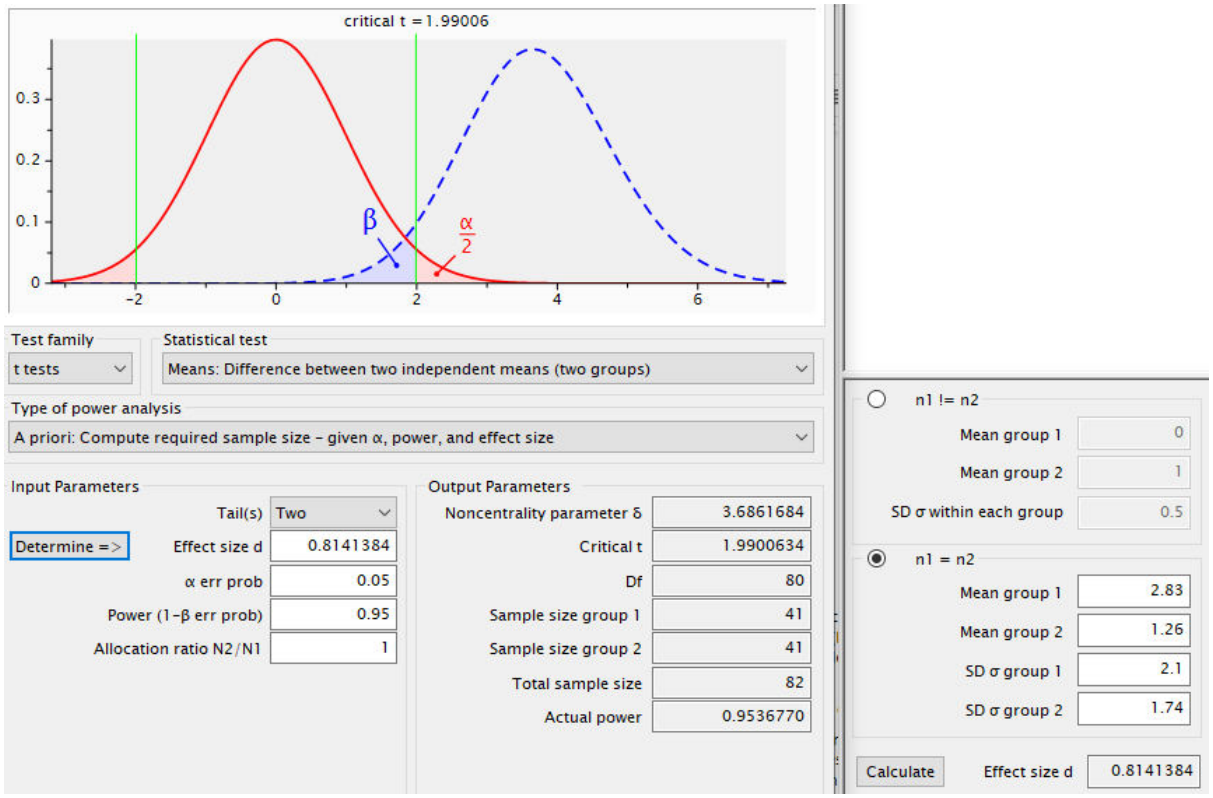


F tests - ANOVA: Fixed effects, special, main effects and interactions
Analysis: A priori: Compute required sample size
Input: Effect size f = 0.2041241
α err prob = 0.05
Power (1-β err prob) = 0.95
Numerator df = 1
Number of groups = 4
Output: Noncentrality parameter λ = 13.0833275
Critical F = 3.8716305
Denominator df = 310
Total sample size = 314
Actual power = 0.9501074

Contrast of Identifiable/no intervention Condition on Donations and the Combination of Other Conditions on Donations (Study 1)

Original result: $t(117) = -4.06$, $p < .001$

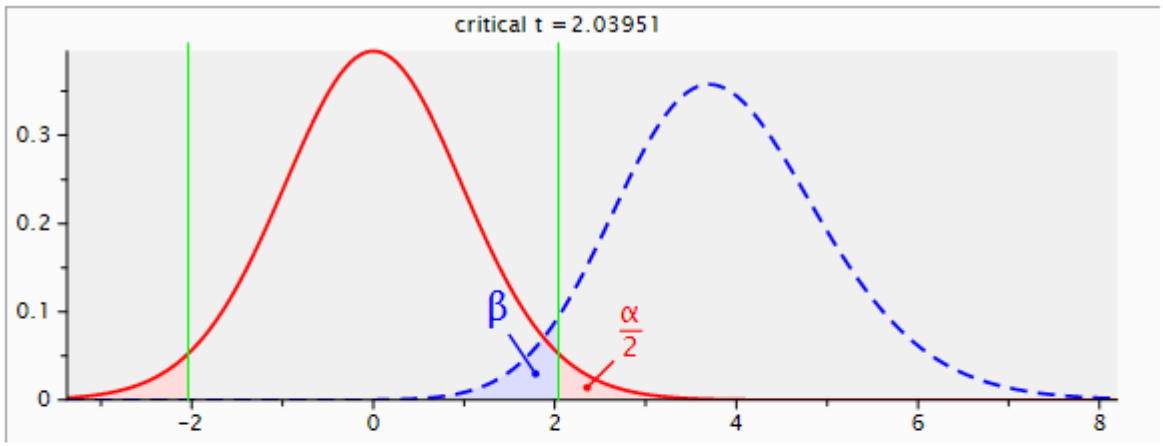
- identifiable/no intervention: M = \$2.83, SD = \$2.10
- Combination of other 3 conditions (statistical/no intervention, statistical /intervention, identifiable/ intervention): M = \$1.26, SD = \$1.74

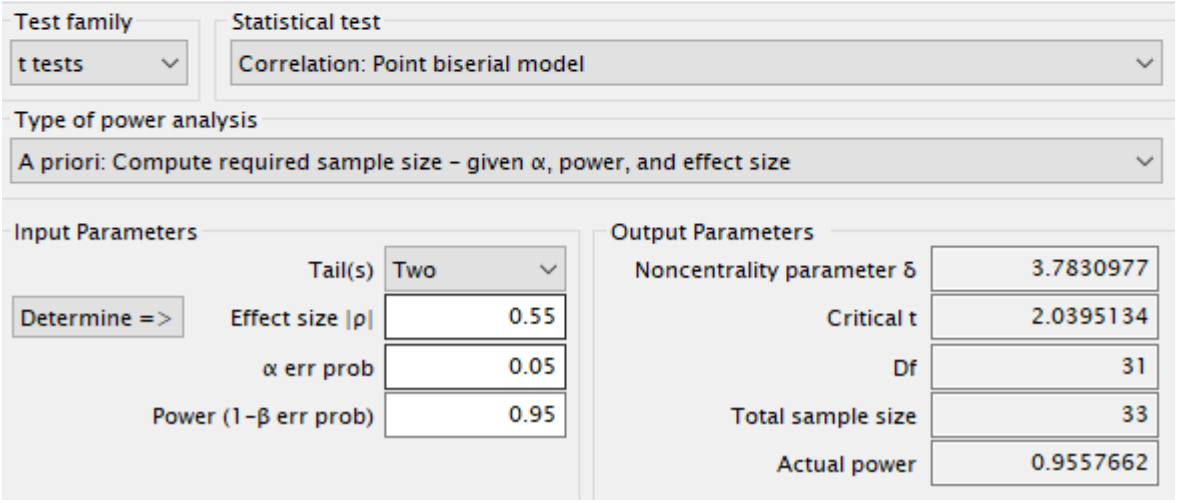


t tests - Means: Difference between two independent means (two groups)
Analysis: A priori: Compute required sample size
Input: Tail(s) = Two
Effect size d = 0.8141384
α err prob = 0.05
Power (1-β err prob) = 0.95
Allocation ratio N2/N1 = 1
Output: Noncentrality parameter δ = 3.6861684
Critical t = 1.9900634
Df = 80
Sample size group 1 = 41
Sample size group 2 = 41
Total sample size = 82
Actual power = 0.9536770

Correlation of Aggregated Feelings and Donations: Identifiable/ no intervention (Study 1)

Original result: $r = .55$, $p < .01$



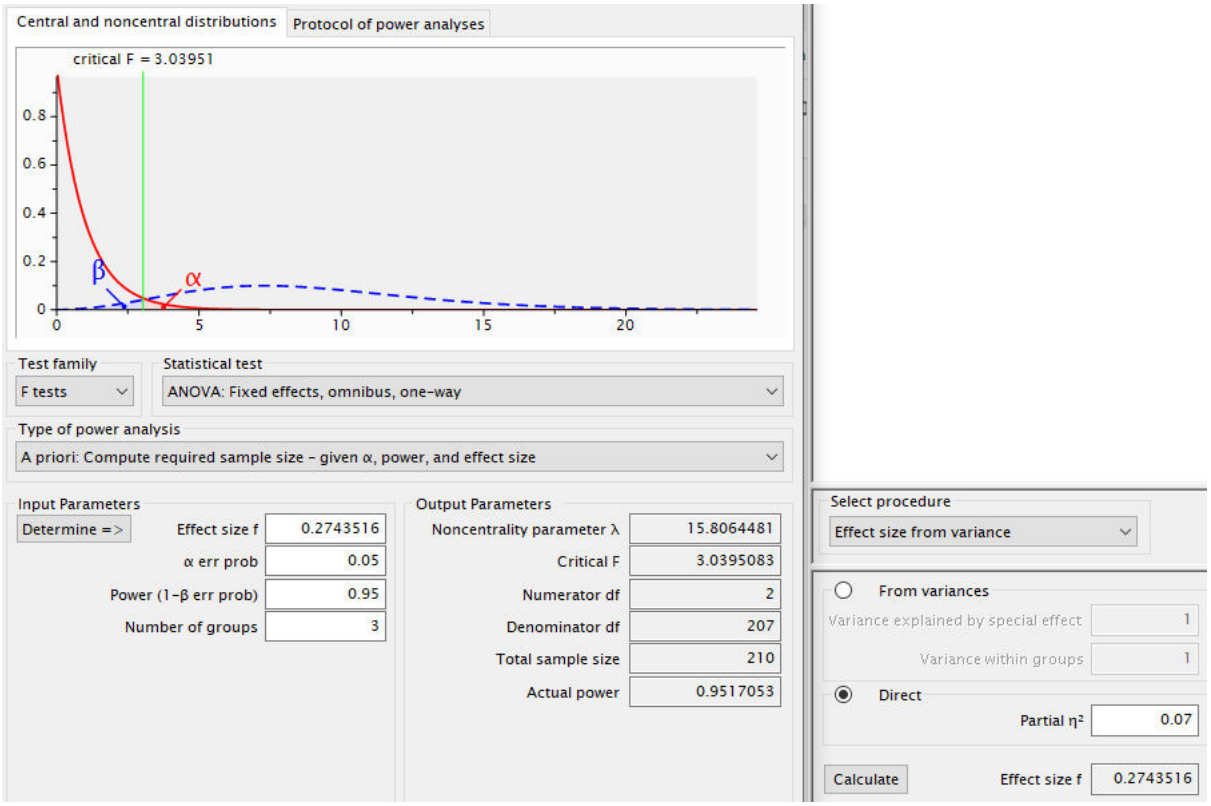


t tests - Correlation: Point biserial model
Analysis: A priori: Compute required sample size
Input: Tail(s) = Two
Effect size |ρ| = 0.55
α err prob = 0.05
Power (1-β err prob) = 0.95
Output: Noncentrality parameter δ = 3.7830977
Critical t = 2.0395134
Df = 31
Total sample size = 33
Actual power = 0.9557662

Main Effect of Identifiability on Donations (Study 3)

Original result: $F(2) = 5.67$, $p < .01$, $\eta_p^2 = .07$

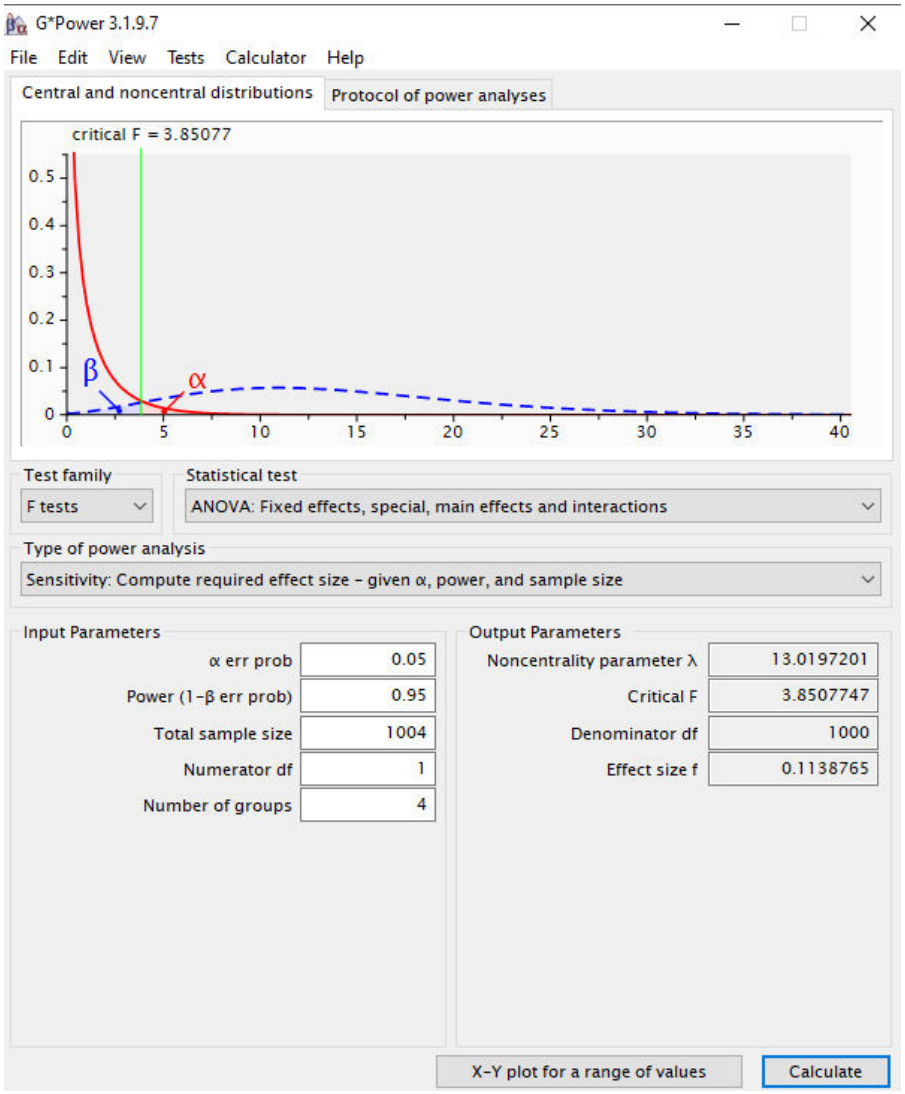
- $\eta_p^2 = .07$ obtained from the original study was directly used for calculating the sample size



F tests - ANOVA: Fixed effects, omnibus, one-way
Analysis: A priori: Compute required sample size
Input: Effect size f = 0.2743516
α err prob = 0.05
Power (1-β err prob) = 0.95
Number of groups = 3
Output: Noncentrality parameter λ = 15.8064481
Critical F = 3.0395083
Numerator df = 2
Denominator df = 207
Total sample size = 210
Actual power = 0.9517053

Below is the summary of the calculated CI, effect size and the required sample size to detect such effect. The maximum required sample size to detect the smallest effect size (the effect of intervention on donations and the interaction of intervention and identifiability on donations) was 314.

Sensitivity Power Analysis



This implies a partial eta squared of 0.012 (partial eta squared = f^2/(1+f^2)

Table S10
Summary of the Confidence Interval, Effect Size and Required Sample Size of the Studied Effects

Effect	**Statistical test**	**Original result**	**Cohens'*d***	**95%CI for Cohen's *d***	**Required sample size**
Main effect of identifiability on donations (Study1)	Two-way ANOVA	$F(1,115) = 6.75$, $p < .05$, $\eta_p^2 = .06$	0.48	[0.11, 0.84]	206
Main effect of intervention on donations (Study1)	Two-way ANOVA	$F(1,115) = 4.15$, $p < .05$,$\eta_p^2 = .04$	-0.36	[-0.72, 0.002]	314
Interaction effect of intervention and identifiability on donations (Study 1)	Two-way ANOVA	$F(1,115) = 5.32$, $p < .05$, $\eta_p^2 = .04$	Partial eta squared was used to calculate the 95%CI	[0.0003, 0.14]	314
Contrast of identifiable/no intervention condition on donations and combination of other conditions on donations	t-test	$t(117) = -4.06$, $p < .001$	-0.85	[-1.27, -0.42]	82
Correlation of aggregated feelings and donations: identifiable/no intervention	Correlation	$r = .55$, $p < .01$	0.55	[0.24, 0.76]	33
Main effect of identifiability on donations (with Joint condition) (Study 3)	One way ANOVA	$F(2) = 5.67$, $p < .01$, $\eta_p^2 = .07$	Partial eta squared was used to calculate the 95%CI	[0.008, 0.15]	210

Materials and Scales Used in the Replication and Extension

Procedure

Table S11 summarizes the procedure adopted in the replication.

Table S11
Procedure of the Replication

Section	Procedure		
Consent of participation	Participants indicated they were 18 or above before giving consent to participate in the survey		
Study outline	Participants were given a brief introduction to the study to confirm their willingness to participate in the survey		
Explicit Learning Intervention (participants were evenly and randomly assigned to two conditions)	Intervention: Read a passage about previous research on the identifiable victim effect	No intervention: *Go straight to the identifiability section*	
Comprehension check	According to the above information, to which type of people we tend to react more strongly? -No difference between an identifiable victim and statistical victims -An identifiable victim -Statistical victim	Not applicable	
Identifiability (participants were evenly and randomly assigned to three conditions)	Identifiable victim condition: - Read a brief description of an African girl displayed in a photo	Statistical victim condition: - Read some factual information about the problems of starvation in Africa	Joint condition: *Victim information was a combination of the materials given in the identifiable and statistical victim conditions*
Comprehension check	According to the above-mentioned descriptions, which of the following countries has the greatest number of children suffering from starvation in Africa? -Uganda -The Democratic Republic of Congo -Ethiopia	Which of the following describes the above-mentioned child? - She is 9 years old - She comes from Ethiopia Her name is Juliet	*Two questions identical to the identifiable and statistical victim condition*

Section	Procedure		
Donations	Imagine that you have just earned $5 US dollars. You are given an opportunity to donate any amount of the money to the organization Save the Children. Money donated will go toward Juliet. Please indicate your choice of donation below. ($0, $1, $2, $3, $4, $5)	Imagine that you have just earned $5 US dollars. You are given an opportunity to donate any amount of the money to the organization Save the Children. Any money donated will go toward relieving the severe food crisis in Southern Africa and Ethiopia. Please indicate your choice of donation below. ($0, $1, $2, $3, $4, $5)	*Identical to the identifiable victim condition*
Feelings	Participants were asked to rate their feelings that best described them at the time of donation: (1 = *not at all*, 5 = *extremely*) 1. Feeling of being upset 2. Feeling of being sympathetic 3. Feeling of being touched 4. Moral responsibility 5. Donation appropriateness 6. Perceived impact of donation (Extension)		
Funneling questions	Five funneling questions were asked: 1. Seriousness in completing the survey 2. Whether they have seen the materials before 3. Purpose of the study 4. Error in the study 5. English proficiency		
Demographics	Six demographic questions were asked: 1. Age 2. Gender 3. Race 4. Country of origin 5. Country of residence 6. Family's social class		
Debriefing	Participants were informed the true purpose of the study and confidentially of the result		

Material Used in the Experiments

Intervention Section

For the intervention group, the following text would be shown to the participants which was identical to the one used in the original Study 1:

> Before the survey begins, we would like to tell you about some research conducted by social scientists. This research shows that **people typically react more strongly to specific people who have problems than to statistics about people with problems**. For example, when ‘‘Baby Jessica’’ fell into a well in Texas in 1989, people sent over $700,000 for her rescue effort. Statistics—e.g., the thousands of children who will almost surely die in automobile accidents this coming year—seldom evoke such strong reactions.

Identifiability Section

The information used in this section was retrieved from the Save the Children website: http://savethechildren.org/

1. Identifiable victim condition

Figure S1. Photo of the little girl (presented in black and white).

A brief description of the little girl:

“Any money that you donated will go to Juliet, a 7-year-old girl from Zambia, Africa. She is desperately poor, and faces a threat of severe hunger or even starvation. Her life will be changed for the better as a result of your financial gift. With your support, and the support of other caring sponsors, Save the Children will work with Juliet’s family and other members of the community to help feed her, provide her with education, as well as basic medical care and hygiene education.”

2. Statistical victim condition

The following text would be shown to the participants:

“Approximately 1.5 million children in Mozambique, Malawi and Zimbabwe have been affected by the devastating impact of Cyclone Idai. Many rescued children and families have not had access to food and water for days.
Across Ethiopia, at least 7.8 million people – including 4 million children – have been hit by drought and are struggling to get enough food and water to feed themselves and their livestock.
There are fears that there will eventually be nearly 1 million child refugees in Uganda. Many children have walked for days without food, water or rest and are in desperate need of healthcare, food and water.
An extreme hunger crisis is threatening some 4.6 million acutely malnourished children in The Democratic Republic of Congo.”

3. Joint condition

Information shown in this condition would be identical to those used in the identifiable victim condition and the statistical victim condition.

Data Analysis Plan

This section summarizes the data analysis plan of the replication.

Main Effect of Identifiability and Intervention on Donations

Table S12
Main Effects of Identifiability and Intervention on Donation

Statistical test	**2x3 two-way ANOVA** - Effect of identifiability on donations - Effect of intervention on donations - Interaction of intervention and identifiability on donations **One way ANOVA** - Effect of identifiability on donations	
Design	Between-subject	
IV	**2x3 two-way ANOVA** Identifiability - identifiable victim condition - statistical victim condition - joint condition **One way ANOVA** Identifiability - identifiable victim condition - statistical victim condition	**2x3 two-way ANOVA** Intervention - no intervention - intervention
DV	Donations	
Note	For the two-way ANOVA, if significant results were found, pairwise independent t-tests would be conducted to indicate which conditions attributed to the significant difference of donations with Bonferroni-adjusted *p*-values reported.	

Note. IV= Independent variable. DV= Dependent variable.

Main Effect of Identifiability and Intervention on Aggregated Feelings

Table S13
Main Effect of Identifiability and Intervention on Aggregated Feelings

Statistical test	2x3 two-way ANOVA - Main effect of identifiability on aggregated feelings - Main effect of intervention on aggregated feelings - Interaction of intervention and identifiability on aggregated feelings	
Design	Between-subject	
IV	Identifiability - identifiable victim condition - statistical victim condition - joint condition	Intervention - no intervention - intervention
DV	Aggregated feelings (by averaging the score of the feelings of being upset, being touched, being sympathetic, moral responsibility, donation appropriateness and perceived impact of donation)	
Note	If significant results were found, pairwise independent t-tests would be conducted to indicate which conditions attributed to the significant difference of aggregated feelings with Bonferroni-adjusted *p*-values reported.	

Note. IV= Independent variable. DV= Dependent variable.

Main effect of identifiability and intervention on segregated feelings

Table S14
Main Effect of Identifiability and Intervention on Segregated Feelings

Statistical test	2x3 two-way ANOVA - Main effect of identifiability on each segregated feeling - Main effect of intervention on each segregated feeling - Interaction of intervention and identifiability on each segregated feeling	
Design	Between-subject	
IV	Identifiability - identifiable victim condition - statistical victim condition - joint condition	Intervention - no intervention - intervention
DV	Segregated feelings: - being upset - being touched - being sympathetic - moral responsibility - donation appropriateness - perceived impact of donation (extension)	
Note	If significant results were found, pairwise independent t-tests would be conducted to indicate which conditions attributed to the significant difference of feelings with Bonferroni-adjusted *p*-values reported.	

Note. IV= Independent variable. DV= Dependent variable.

Correlation between Aggregated Feelings and Donations Across Conditions

Table S15
Correlations between Aggregated Feelings and Donations

Main effect	Correlation between aggregated feelings and donations
Statistical test	Linear Pearson correlation between aggregated feelings and donations across the following conditions - identifiable/no intervention - statistical/no intervention - joint/no intervention - identifiable/intervention - statistical/intervention - joint/intervention
Design	Between-subject
Predictor	Aggregated feelings (by averaging the feelings of being upset, being touched, being sympathetic, moral responsibility, donation appropriateness and perceived impact of donation)
Outcome	Donations

Note. IV= Independent variable. DV= Dependent variable.

Comparisons and Deviations

We made many changes to the wording of the hypotheses in comparison to the preregistration to improve the formulation of interaction predictions following peer review. The main changes had to do with a more careful mapping of the hypotheses to the target article and differentiating between a direct replication using the same conditions as in the target article's studies, and extension analyses that were made possible because of the unified design.

Original versus Pre-registration versus Replication: General Summary of Deviation

Table S16
Summary of deviations

	Original Study	Pre-registration	Replication	Reason for change
Sample				
Sample nature	People in the student center at a university in Pennsylvania	MTurk workers	MTurk workers	To generalize the result with participants coming from a more diverse background
Sample size	Study 1: 121 Study 3: 159	1000	1004	The sample size for replication was adjusted to meet the funding requirement of compensation in this project
Physical setting	Student center at a university in Pennsylvania	Qualtrics	Qualtrics	Convenience of data collection
Mode of data collection	Pen and paper	Online	Online	Convenience of data collection
Survey design				
Pre-survey	A pre-survey about the use of various technological products was conducted to allow participants to receive $5 for donation purpose	No pre-survey	No pre-survey	Imaginary donation was adopted instead of real time donation
Display format	Separated the identifiable, statistical victim and joint condition into two studies	Combined the identifiable, statistical victim and joint condition into a single study	Combined the identifiable, statistical victim and joint condition into a single study	To combine two studies into a single study for comparison
Comprehension check	Not specified	Questions would be displayed after participants read the victim information and research about the phenomenon	Questions would be displayed after participants read the victim information and research about the phenomenon	To make sure the participants read the information before they proceed to the next question
Intervention	Only applied to the identifiable and statistical victim condition	Applied to the joint condition as well	Applied to the joint condition as well	To allow comparison of the joint condition with or without intervention
Affective feelings	Only applied to the identifiable and statistical victim condition	Applied to the joint condition as well	Applied to the joint condition as well	To study how the affective feelings differed in the joint condition
Donation procedure	Filled in the charity request letter and included the donated money in an envelope which would be donated to Save the Children	Imaginary donation	Imaginary donation	To make the donation procedure more compatible in an online setting

Original versus Replication: Identifiability Scenario

Table S17
Summary of Deviation for the Identifiability Scenario

Scenario	Original	Replication	Reason for change
Statistical victim condition	Food shortages in Malawi are affecting more than 3 million. In Zambia, severe rainfall deficits have resulted in a 42 percent drop in maize production from 2000. As a result, an estimated three million Zambians face hunger. Four million Angolans- one third of the population- have been forced to flee their homes. More than 11 million people in Ethiopia need immediate food assistance.	Approximately 1.5 million children in Mozambique, Malawi and Zimbabwe have been affected by the devastating impact of Cyclone Idai. Many rescued children and families have not had access to food and water for days. Across Ethiopia, at least 7.8 million people – including 4 million children – have been hit by drought and are struggling to get enough food and water to feed themselves and their livestock. There are fears that there will eventually be nearly 1 million child refugees in Uganda. Many children have walked for days without food, water or rest and are in desperate need of healthcare, food and water. An extreme hunger crisis is threatening some 4.6 million acutely malnourished children in the Democratic Republic of Congo .	The most updated statistical victim information was used given that the statistics in the original study was collected around the same period of time the experiment was conducted
Identifiable victim condition	"Any money that you donate will go to **Rokia**, a 7-year-old girl from **Mali**, Africa. **Rokia** is desperately poor, and faces a threat of severe hunger or even starvation. Her life will be changed for the better as a result of your financial gift. With your support, and the support of other caring sponsors, Save the Children will work with **Rokia**'s family and other members of the community to help feed her, provide her with education, as well as basic medical care and hygiene education."	"Any money that you donate will go to **Juliet**, a 7-year-old girl from **Zambia**, Africa. **Juliet** is desperately poor, and faces a threat of severe hunger or even starvation. Her life will be changed for the better as a result of your financial gift. With your support, and the support of other caring sponsors, Save the Children will work with **Juliet**'s family and other members of the community to help feed her, provide her with education, as well as basic medical care and hygiene education."	The photo of Rokia was not available in the original study which was replaced by Juliet found in the website of Save the Children

Note. The differences between the original study and replication were bolded.

Original versus Replication: Questions Asking about Affective Feelings

Table S18
Summary of Deviation for Affective Feelings

	Original	Replication	Reason for change
Being upset	How upsetting is this **situation** to you?	How upsetting is the described **situation of the victim(s)** to you?	To make sure the reported emotion was provoked from the situation describing the victim(s)
Being sympathetic	How sympathetic did you feel while reading the description of the **cause**?	How sympathetic did you feel while reading the description of the **victim(s)**?	
Being touched	How touched were you by the **situation described**?	How touched were you by the described **situation of the victim(s)**?	
Moral responsibility	How much do you feel it is your moral responsibility to help out with this **cause**?	How much do you feel it is your moral responsibility to help out with the **victim(s)**?	
Donation appropriateness	To what extent do you feel that it is appropriate to give money to aid this **cause**?	To what extent do you feel that it is appropriate to give money to aid the **victim(s)**?	

Original versus Replication: Questions Asking for Donations

Table S19
Summary of Deviation for Donations

Donation target	Original	Replication	Reason for change
An identifiable victim	Now that you have had the opportunity to learn about how any money you donate will be used, please fill out the following page and include it with any money you donate in the envelope you have been given. Even if you do not choose to donate, please fill out the form and return it to us in the envelope.	Imagine that you've just earned $5 US dollars. You're given an opportunity to donate any amount of the money to the organization Save the Children. Money donated will go toward Juliet.	To specify the money they donated would go to corresponding victims
Statistical victims	Same as above	Imagine that you have just earned $5 US dollars and you are given an opportunity to donate any amount of the money to the organization Save the Children. Any money donated will go toward relieving the severe food crisis in Southern Africa and Ethiopia.	To specify the money they donated would go to corresponding victims

Pre-exclusions versus Post-exclusions

Table S20 and S21 summarize the number of cases for exclusion.

Table S20
Summary of Pre-exclusion and Post-exclusion

	Before exclusion	Cases fulfilling exclusion criteria	After exclusion
Number of cases	1004	56	948

Table S21
Summary of Exclusion Cases

Exclusion criteria	Cases fulfilling exclusion criteria
Participants indicating a low proficiency of English (Exclusion: self-report < 4, on a 0-6 scale)	5
Participants who self-report not being serious about filling in the survey (Exclusion: self-report 3, on a 0-4 scale)	19
Participants who correctly guessed the hypothesis of this study in the funneling section. (Exclusion: include the word " identifiable victim effect")	0
Participants who have already seen or done the survey before. (Exclusion: answered "yes")	45
Participants who failed to complete the survey. (Exclusion: duration = 0, leave question blank)	0
Participants not from the United States	0

Note. Some participants are excluded for multiple reasons; therefore, the sum over these exclusion criteria is more than the total number of exclusions.

Deviations of Replication from the Pre-registered Plan

Table S22
Summary of Deviation between Pre-registration and Replication

	Pre-registration	Replication	Reasons for deviation
Handling outliers	Adopted the method of median absolute deviation (MAD) in detecting univariate outliers	Did not detect outliers	We realized that it is less relevant to handle outliers with MAD given that our scale used in donation was fixed in a range from $0 to $5 and that the options were predetermined by researchers
Assumption of normality	Was not considered	Reported the skewness and kurtosis of dependent variables	We realized that central limit theorem would ensure normality of sample means even if the data is not normally distributed.
Data analysis	Considered all three conditions of identifiable, statistical victim, and joint conditions in the main analysis for the main effects of identifiability, intervention and their interaction on donations	Separated the main effects of identifiability, intervention, and their interaction with or without joint conditions	Allowed a fair comparison with the results of the original studies.

Additional Analyses and Results

Comparison of Data Analysis Between Original Study And Replication

Table S23
Comparison of Data Analysis Between the Original Study and Replication

Hypothetical effect	Statistical testing	Hypothetical effect	
		Original study	Replication
Main effect: Identifiability and intervention on donations	Two way ANOVA	Effect of identifiability on donations **(without joint condition)**	Identical to the original study
		Effect of intervention on donations **(without joint condition)**	
		Interaction of identifiability and intervention on donations **(without joint condition)**	
	Two way ANOVA	N/A	Effect of identifiability on donations **(with joint condition)**
			Effect of intervention on donations **(with joint condition)**
			Interaction of identifiability and intervention on donations **(with joint condition)**
	One way ANOVA	Effect of identifiability on donations **(with joint condition)**	Embedded into the two way ANOVA illustrated above

	Post-hoc contrast test (Note: we conduct the test only if the interaction effect was significant)	Contrast the mean difference of identifiable/no intervention condition on donations with the **combination of 3 conditions** (statistical/no intervention, statistical/intervention, identifiable/intervention)	**We did not conduct this test in this replication given the insignificant results for the interaction effect**
	Ordered probit regression	Effect of identifiability on donations Effect of intervention on donations Interaction of identifiability and intervention on donations	**We did not conduct this test in this replication**
Identifiability and intervention on aggregated feelings	Two way ANOVA	Effect of identifiability on aggregated feelings **(without joint condition)** Effect of intervention on aggregated feelings **(without joint condition)** Interaction of identifiability and intervention on aggregated feelings **(without joint condition)**	Effect of identifiability on aggregated feelings **(with joint condition)** Effect of intervention on aggregated feelings **(with joint condition)** Interaction of identifiability and intervention on aggregated feelings **(with joint condition)**
Identifiability and intervention on segregated feelings	Two way ANOVA	Effect of identifiability on segregated feelings **(without joint condition)** Effect of intervention on segregated feelings **(without joint condition)** Interaction of identifiability and	Effect of identifiability on segregated feelings **(with joint condition)** Effect of intervention on segregated feelings **(with joint condition)** Interaction of identifiability and

		intervention on segregated feelings **(without joint condition)**	intervention on segregated feelings **(with joint condition)**
Correlation between aggregated feelings and donations	Linear Pearson's correlation	Correlation between aggregated feelings and donations across **four different conditions** (statistical/intervention, statistical/no intervention, identifiable/intervention, identifiable/no intervention)	Correlation between aggregated feelings and donations across **six different conditions** (statistical/intervention, statistical/no intervention, identifiable/intervention, identifiable/no intervention, **joint/intervention, joint/no intervention**)

Note. Bolded items denote the differences between the original and replication study. Segregated feelings include the feelings of being upset, being sympathetic, being touched, moral responsibility, donation appropriateness. Aggregated feelings refer to averaging all the segregated feelings into a single composite. "Perceived impact of donation" was also included in the replication in the calculation of segregated and aggregated feelings.

Evaluation of Replication 'Closeness'

Table S25 summarizes the current replication as a "close to very close replication". For more details about the deviation from the original study, please refer to the supplementary session of "Comparisons and deviations".

Table S25
Criteria for Evaluation of Replications by Lebel et al. (2018).

Target similarity	Highly similar				Highly dissimilar
Category	Direct replication			Conceptual replication	
Design facet	**Exact replication**	**Very close replication**	**Close replication**	**Far replication**	**Very far replication**
IV operationalization	Same/similar	Same/similar	Same/similar	Different	
DV operationalization	Same/similar	Same/similar	Same/similar	Different	
IV stimuli	Same/similar	Same/similar	Different		
DV stimuli	Same/similar	Same/similar	Different		
Procedural details	Same/similar	Different			
Physical setting	Same/similar	Different			
Contextual variables	Different				

Note. A classification of relative methodological similarity of a replication study to an original study. "Same" ("different") indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable. "Everything controllable" indicates design facets over which a researcher has control. Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.).
"Similar" category was added to the Lebel et al. (2018) typology to refer to minor deviations aimed to adjust the study to the target sample that is not expected to have major implications on replication success.

Descriptive Statistics and Statistical Tests for Segregated Feelings

Table S26
Identifiability on Segregated Feelings: Descriptive Statistics

Segregated feelings	Identifiable victim condition		Statistical victim condition		Joint condition	
	M	*SD*	*M*	*SD*	*M*	*SD*
Being upset	3.68	.10	3.84	1.18	3.77	1.09
Being sympathetic	4.03	.00	3.99	1.10	3.84	1.14
Being touched	3.75	.06	3.73	1.18	3.63	1.23
Moral responsibility	3.47	.24	3.57	1.26	3.30	1.35
Donation appropriateness	4.13	.00	4.05	1.14	3.90	1.17
Perceived impact of donation (Extension)	3.39	.24	2.96	1.35	3.02	1.39

Note. $N = 335$ for identifiable victim condition. $N = 335$ for statistical victim condition. $N = 334$ for joint condition.

Table S27
Segregated Feelings: Statistical Tests

Variations	*df*	*F*	η_p^2	95% CI	*p*	BF_{01}
Identifiability						
Being upsetting	2	1.53	.003	[0.000, 0.012]	.22	18.35
Being sympathetic	2	2.82	.006	[0.000, 0.017]	.060	4.94
Being touched	2	0.90	.002	[0.000, 0.009]	.41	33.12
Moral responsibility	2	3.82	.008	[0.000, 0.021]	.022	2.05
Donation appropriateness	2	3.71	.007	[0.000, 0.020]	.025	2.24
Perceived impact (Extension)	2	10.48	.021	[0.006, 0.040]	<.001	0.003
Intervention						
Being upset	1	2.17	.002	[0.000, 0.012]	.14	4.55
Being sympathetic	1	1.84	.002	[0.000, 0.011]	.18	5.41
Being touched	1	0.72	.0007	[0.000, 0.008]	.40	9.75
Moral responsibility	1	0.046	.00005	[0.000, 0.004]	.83	14.05
Donation appropriateness	1	0.67	.001	[0.000, 0.008]	.41	9.87
Perceived impact (Extension)	1	0.11	.0001	[0.000, 0.005]	.74	13.69
Interaction						
Being upset	2	0.024	.00005	[0.000, 1.000]	.98	45.53
Being sympathetic	2	0.81	.002	[0.000, 0.009]	.44	22.17
Being touched	2	1.98	.004	[0.000, 0.014]	.14	6.36
Moral responsibility	2	0.67	.001	[0.000, 0.008]	.51	21.66
Donation appropriateness	2	1.11	.002	[0.000, 0.010]	.33	14.49
Perceived impact (Extension)	2	1.48	.003	[0.000, 0.012]	.23	11.43
Residual	998					

Note. ANOVA test. $N = 1004$. CI = confidence interval.

Summary of Pairwise Independent t-tests for Perceived Impact of Donation

Table S28
Perceived Impact of Donation: Pairwise Independent t-tests

Comparison							
Identifiability	Identifiability	*df*	*t*	*p*	BF_{01}	Cohen's *d*	95% CI
Statistical victim	Identifiable victim	668	-4.35	.00005	0.001	-0.34	[-0.49, -0.18]
Joint condition	Identifiable victim	667	-3.67	.0008	0.02	-0.28	[-0.44, -0.13]
Statistical victim	Joint condition	667	-0.59	1.00	9.76	-0.046	[-0.20, 0.11]

Note. Independent t- test. $N = 335$ for identifiable victim condition. $N = 335$ for statistical victim condition. $N = 334$ for joint condition. CI = confidence interval. *p* values reported were Bonferroni-adjusted.

Table S29
Correlations between Aggregated Feelings and Hypothetical donations across Conditions

	N	*r*	95% CI	*p*
Identifiable/ Explicit learning intervention	170	.53	[0.41, 0.63	< .001
Identifiable/ No explicit learning intervention	165	.48	[0.35, 0.59]	< .001
Statistical/ Explicit learning intervention	159	.60	[0.49, 0.69]	< .001
Statistical/No explicit learning intervention	176	.49	[0.37, 0.59]	< .001
Joint/ Explicit learning intervention	173	.57	[0.46, 0.66]	< .001
Joint/ No explicit learning intervention	161	.59	[0.48, 0.68]	< .001

Exclusion Based on Pre-registration Criteria

A total of 1004 samples were collected from Amazon MTurk Platform. According to the exclusion criteria set out in the pre-registration, our main reporting was on the full sample of participants who finished the study, yet we ran an additional analyses using additional criteria resulting in 948 participants. The results after exclusion were similar to those before exclusion regarding the significance of different statistical tests.

Table S30
Identifiability and Intervention on Donations: Statistical Tests

	F	*p*	BF_{01}	η_p^2	95% CI
Identifiability (without joint condition) [Study 1]					
Original study	6.75	< .05		.06	[.00, .15]
Replication	0.16	0.699	10.51	.0002	[.000, .008]
Identifiability (with joint condition) [Study 3]					
Original study	5.67	< .01		.07	[.01, .15]
Replication	3.26	.039	3.03	.007	[.000, .020]
Identifiability (with joint condition & without explicit learning)	0.60	.55	24.04	.002	[.000, .016]
Intervention (without joint condition) [Study 1]					
Original study	4.15	< .05		.03	[.00, .12]
Replication	0.16	0.692	10.5	.0002	[.000, .008]
Intervention (With joint condition) [Study 3]					
Extension	0.37	.542	11.00	.0003	[.000, .006]
Interaction (without joint condition) [Study 1]					
Original study	5.32	< .05		.04	[.00, .14]
Replication	0.654	.431	5.90	.001	[.000, .012]
Interaction (with joint condition) [Study 3]					
Extension	1.49	.227	10.41	.003	[.000, .013]

Note. ANOVA test. $N = 1004$. CI = confidence interval. N/A = Unreported or not relevant in the study identifiable victim, statistical victim, and joint conditions were considered. For explicit learning, only the identifiable victim and statistical victim conditions were considered. BF_{01} denotes the Bayes factor in favor of the null.

Table S31
Identifiability on Donations (With Joint Condition): Pairwise Independent t-tests

identifiability	identifiability	t	df	p-bonf	Cohen's d	Lower	Upper
IDV	IDV&SV	2.0594948	626.0000	0.1195795	0.16436641	0.007376039	0.32135679
	SV	-0.3847058	633.0000	1.0000000	-0.03053411	-0.186403625	0.12533540
IDV&SV	SV	-2.4030304	631.0000	0.0496425	-0.19103545	-0.347503194	-0.03456770

Note. *p* values reported were Bonferroni-adjusted.

Bayesian analysis

IDV vs IDV & SV -> BF01 = 1.42

IDV vs SV -> BF01 = 10.51

IDV & SV vs SV -> BF01 = 0.68

Table S32
Identifiability and Intervention on Aggregated Feelings: Statistical Tests

A: With Joint Condition

	Sum of Squares	df	Mean Square	F	p	η*p
Intervention	1.946447	1	1.9464471	2.0249303	0.1550671	0.0021450
identifiability	3.379406	2	1.6897030	1.7578340	0.1729830	0.0037183
Intervention:identifiability	1.471278	2	0.7356389	0.7653008	0.4654830	0.0016222
Residuals	905.489512	942	0.9612415			

B: Without Joint Condition

	Sum of Squares	df	Mean Square	F	p	η*p
Intervention	0.20432365	1	0.20432365	0.21888817	0.6400499	0.0003468
identifiability	0.34345711	1	0.34345711	0.36793929	0.5443478	0.0005828
Intervention:identifiability	0.05940510	1	0.05940510	0.06363959	0.8009154	0.0001008
Residuals	589.01411810	631	0.93346136			

Note. Cronbach's alpha with exclusions was 0.90

Bayesian analysis with joint condition

Identifiability -> BF01 = 13.31
Intervention -> BF01 = 4.71
Interaction -> BF01 = 21.36

Bayesian analysis without joint condition

Identifiability -> BF01 = 9.32
Intervention -> BF01 = 10.04
Interaction -> BF01 = 7.76

Table S33
Feeling of Being Upset

	Sum of Squares	df	Mean Square	F	p	η²p
identifiability	5.3049306	2	2.6524653	2.08823266	0.1244791	0.0044140
Intervention	3.8666546	1	3.8666546	3.04413948	0.0813552	0.0032212
identifiability:Intervention	0.2456569	2	0.1228285	0.09670038	0.9078370	0.0002053
Residuals	1196.5248809	942	1.2701963			

Bayesian analysis

Identifiability -> BF01 = 9.80
Intervention -> BF01 = 2.79
Interaction -> BF01 = 39.53

Table S34
Feeling of Being Sympathetic

	Sum of Squares	df	Mean Square	F	p	η²p
identifiability	4.570461	2	2.285230	1.948654	0.1430396	0.0041202
Intervention	3.264902	1	3.264902	2.784036	0.0955403	0.0029467
identifiability:Intervention	2.519489	2	1.259745	1.074205	0.3419872	0.0022755
Residuals	1104.704600	942	1.172723			

Bayesian analysis

Identifiability -> BF01 = 10.82
Intervention -> BF01 = 3.29
Interaction -> BF01 = 15.45

Table S35
Feeling of Being Touched

	Sum of Squares	df	Mean Square	F	p	η²p
identifiability	1.896405	2	0.9482026	0.6938891	0.4998844	0.0014711
Intervention	2.307632	1	2.3076321	1.6887117	0.1940888	0.0017895
identifiability:Intervention	5.242377	2	2.6211885	1.9181704	0.1474487	0.0040560
Residuals	1287.247236	942	1.3665045			

Bayesian analysis

Identifiability -> BF01 = 38.32
Intervention -> BF01 = 5.87
Interaction -> BF01 = 7.67

Table S36
Moral Responsibility

	Sum of Squares	df	Mean Square	F	p	η*p
identifiability	12.8963367	2	6.4481684	3.87154606	0.0211583	0.0081528
Intervention	0.1134182	1	0.1134182	0.06809744	0.7941844	0.0000723
identifiability:Intervention	2.0636144	2	1.0318072	0.61950756	0.5384285	0.0013136
Residuals	1568.9273759	942	1.6655280			

Bayesian analysis

Identifiability -> BF01 = 1.80
Intervention -> BF01 = 12.96
Interaction -> BF01 = 21.78

Table S37
Donation Appropriateness

	Sum of Squares	df	Mean Square	F	p	η*p
identifiability	5.712895	2	2.856447	2.3462681	0.0962849	0.0049568
Intervention	1.813348	1	1.813348	1.4894725	0.2226037	0.0015787
identifiability:Intervention	2.016997	2	1.008499	0.8283745	0.4370766	0.0017557
Residuals	1146.831193	942	1.217443			

Bayesian analysis

	denominator				
numerator	identifiability	Intervention	identifiability + Intervention	identifiability + Intervention + identifiability:Intervention	
Intercept only	0.01104356	8.843564	0.0902643	0.9259169	

Bayesian analysis

Identifiability -> BF01 = 8.07
Intervention -> BF01 = 6.48
Interaction -> BF01 = 19.30

Table S38
Perceived Impact of Donation

	Sum of Squares	df	Mean Square	F	p	η*p
identifiability	32.845519	2	16.422759	9.261822	0.0001039	0.0192849
Intervention	1.930436	1	1.930436	1.088694	0.2970285	0.0011544
identifiability:Intervention	5.465477	2	2.732738	1.541162	0.2146715	0.0032614
Residuals	1670.323599	942	1.773167			

Bayesian analysis

Identifiability -> BF01 = 0.01
Intervention -> BF01 = 8.84
Interaction -> BF01 = 9.77

Table S39
Perceived Impact of Donation: Pairwise Independent t-tests

identifiability	identifiability	t	df	p-bonf	Cohen's d	Lower	Upper
IDV	IDV&SV	3.3438119	626.0000	0.0026270	0.26686660	0.1094446	0.4242886
	SV	4.1439798	633.0000	0.0001163	0.32890784	0.1719972	0.4858185
IDV&SV	SV	0.6881891	631.0000	1.0000000	0.05470947	-0.1014318	0.2108508

Note. Bonferroni corrected p-value

Bayesian analysis

IDV vs IDV & SV -> BF01 = 0.05

IDV vs SV -> BF01 = 0.002

IDV & SV vs SV -> BF01 = 8.95

Table S40
Correlation between Aggregated Feelings and Donations (Identifiable victim/ no intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.5795138	–
	p-value	< .0000001	–
	95% CI Upper	0.6757388	–
	95% CI Lower	0.4638697	–
	N	154	–

Table S41
Correlation between Aggregated Feelings and Donations (Statistical victim/ no intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.5517379	–
	p-value	< .0000001	–
	95% CI Upper	0.6489365	–
	95% CI Lower	0.4368193	–
	N	168	–

Table S42
Correlation between Aggregated Feelings and Donations (Joint Condition/ No intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.6033475	–
	p-value	< .0000001	–
	95% CI Upper	0.6960011	–
	95% CI Lower	0.4909309	–
	N	151	–

Table S43
Correlation between Aggregated Feelings and Donations (Identifiable victim/ Intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.6279452	–
	p-value	< .0000001	–
	95% CI Upper	0.7133357	–
	95% CI Lower	0.5241828	–
	N	161	–

Table S44
Correlation between Aggregated Feelings and Donations (Statistical victim/ Intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.6436821	–
	p-value	< .0000001	–
	95% CI Upper	0.7282551	–
	95% CI Lower	0.5397969	–
	N	152	–

Table S45
Correlation between Aggregated Feelings and Donations (Joint condition/ Intervention)

		Combined.donation	Feelings
Combined.donation	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Feelings	Pearson's r	0.6379104	–
	p-value	< .0000001	–
	95% CI Upper	0.7211695	–
	95% CI Lower	0.5364852	–
	N	162	–

Table S46
Correlation between Perceived Impact and Donations (Identifiable victim/ no intervention)

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.4671946	–
	p-value	< .0000001	–
	95% CI Upper	0.5823262	–
	95% CI Lower	0.3336910	–
	N	154	–

Table S47
Correlation between Perceived Impact and Donations (Statistical victim/ no intervention)

Correlation Matrix

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.4881814	–
	p-value	< .0000001	–
	95% CI Upper	0.5955694	–
	95% CI Lower	0.3636510	–
	N	168	–

Table S48
Correlation between Perceived Impact and Donations (Joint Condition/ No intervention)

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.5946375	–
	p-value	< .0000001	–
	95% CI Upper	0.6889309	–
	95% CI Lower	0.4805521	–
	N	151	–

Table S49
Correlation between Perceived Impact and Donations (Identifiable victim/ Intervention)

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.5108272	–
	p-value	< .0000001	–
	95% CI Upper	0.6167698	–
	95% CI Lower	0.3867069	–
	N	161	–

Table S50
Correlation between Perceived Impact and Donations (Statistical victim/ Intervention)

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.5922525	–
	p-value	< .0000001	–
	95% CI Upper	0.6867058	–
	95% CI Lower	0.4781336	–
	N	152	–

Table S51
Correlation between Perceived Impact and Donations (Joint condition/ Intervention)

		Donations	Combined_impact
Donations	Pearson's r	–	
	p-value	–	
	95% CI Upper	–	
	95% CI Lower	–	
	N	–	
Combined_impact	Pearson's r	0.5763033	–
	p-value	< .0000001	–
	95% CI Upper	0.6708820	–
	95% CI Lower	0.4632761	–
	N	162	–

Individual RoBMA Models

Table S52
Summary of the Individual RoBMA models.

Robust Bayesian meta-analysis
Models overview:

Model	Prior Effect	Prior Heterogeneity	Prior Bias	Prior prob.	log(marglik)	Post. prob.	Inclusion BF
1	S(0)	S(0)		0.125	-3.26	0.000	0.000
2	S(0)	S(0)	omega[2s: .05] ~ CumD(1, 1)	0.010	7.32	0.055	5.476
3	S(0)	S(0)	omega[2s: .1, .05] ~ CumD(1, 1, 1)	0.010	8.68	0.212	25.490
4	S(0)	S(0)	omega[1s: .05] ~ CumD(1, 1)	0.010	7.27	0.052	5.160
5	S(0)	S(0)	omega[1s: .05, .025] ~ CumD(1, 1, 1)	0.010	8.67	0.210	25.200
6	S(0)	S(0)	omega[1s: .5, .05] ~ CumD(1, 1, 1)	0.010	4.35	0.003	0.264
7	S(0)	S(0)	omega[1s: .5, .05, .025] ~ CumD(1, 1, 1, 1)	0.010	5.86	0.013	1.214
8	S(0)	S(0)	PET ~ C(0, 1)[0, Inf]	0.031	-0.58	0.000	0.002
9	S(0)	S(0)	PEESE ~ C(0, 5)[0, Inf]	0.031	-1.59	0.000	0.001
10	S(0)	Ig(1, 0.15)		0.125	5.21	0.079	0.601
11	S(0)	Ig(1, 0.15)	omega[2s: .05] ~ CumD(1, 1)	0.010	6.06	0.015	1.488
12	S(0)	Ig(1, 0.15)	omega[2s: .1, .05] ~ CumD(1, 1, 1)	0.010	6.57	0.026	2.516
13	S(0)	Ig(1, 0.15)	omega[1s: .05] ~ CumD(1, 1)	0.010	7.75	0.083	8.624
14	S(0)	Ig(1, 0.15)	omega[1s: .05, .025] ~ CumD(1, 1, 1)	0.010	8.06	0.114	12.166
15	S(0)	Ig(1, 0.15)	omega[1s: .5, .05] ~ CumD(1, 1, 1)	0.010	5.91	0.013	1.278
16	S(0)	Ig(1, 0.15)	omega[1s: .5, .05, .025] ~ CumD(1, 1, 1, 1)	0.010	6.33	0.020	1.950
17	S(0)	Ig(1, 0.15)	PET ~ C(0, 1)[0, Inf]	0.031	5.52	0.027	0.863
18	S(0)	Ig(1, 0.15)	PEESE ~ C(0, 5)[0, Inf]	0.031	5.07	0.017	0.543
19	N(0, 1)	S(0)		0.125	-5.91	0.000	0.000
20	N(0, 1)	S(0)	omega[2s: .05] ~ CumD(1, 1)	0.010	3.79	0.002	0.151
21	N(0, 1)	S(0)	omega[2s: .1, .05] ~ CumD(1, 1, 1)	0.010	5.11	0.006	0.567
22	N(0, 1)	S(0)	omega[1s: .05] ~ CumD(1, 1)	0.010	3.06	0.001	0.073
23	N(0, 1)	S(0)	omega[1s: .05, .025] ~ CumD(1, 1, 1)	0.010	4.44	0.003	0.291
24	N(0, 1)	S(0)	omega[1s: .5, .05] ~ CumD(1, 1, 1)	0.010	0.17	0.000	0.004
25	N(0, 1)	S(0)	omega[1s: .5, .05, .025] ~ CumD(1, 1, 1, 1)	0.010	1.66	0.000	0.018
26	N(0, 1)	S(0)	PET ~ C(0, 1)[0, Inf]	0.031	-4.53	0.000	0.000
27	N(0, 1)	S(0)	PEESE ~ C(0, 5)[0, Inf]	0.031	-5.36	0.000	0.000
28	N(0, 1)	Ig(1, 0.15)		0.125	4.00	0.024	0.169
29	N(0, 1)	Ig(1, 0.15)	omega[2s: .05] ~ CumD(1, 1)	0.010	4.53	0.003	0.319
30	N(0, 1)	Ig(1, 0.15)	omega[2s: .1, .05] ~ CumD(1, 1, 1)	0.010	4.93	0.005	0.477
31	N(0, 1)	Ig(1, 0.15)	omega[1s: .05] ~ CumD(1, 1)	0.010	5.04	0.006	0.534
32	N(0, 1)	Ig(1, 0.15)	omega[1s: .05, .025] ~ CumD(1, 1, 1)	0.010	5.24	0.007	0.653
33	N(0, 1)	Ig(1, 0.15)	omega[1s: .5, .05] ~ CumD(1, 1, 1)	0.010	3.26	0.001	0.090
34	N(0, 1)	Ig(1, 0.15)	omega[1s: .5, .05, .025] ~ CumD(1, 1, 1, 1)	0.010	3.59	0.001	0.125
35	N(0, 1)	Ig(1, 0.15)	PET ~ C(0, 1)[0, Inf]	0.031	3.27	0.003	0.088
36	N(0, 1)	Ig(1, 0.15)	PEESE ~ C(0, 5)[0, Inf]	0.031	2.74	0.002	0.052

Note. For more information on the individual models see Bartos et al., (2022)

References

George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson.

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational behaviour and Human Decision Processes, 102*(2), 143–153. doi:10.1016/j.obhdp.2006.01.005

Gilad Feldman

From: Andreas Gloeckner <andreas.gloeckner@uni-koeln.de>
Sent: 28 November 2022 14:44
To: Gilad Feldman
Cc: 'Maximilian Maier'
Subject: AW: [Follow-up request to forward decision-letters to other journals]: Action Letter JDM[220823]-R1

Follow Up Flag: Follow up
Flag Status: Completed

Dear Dr. Feldman:

Both reviewers agreed that the reviews can be reused and I agree concerning the action letter.
Hence, you can reuse the action letter and reviews for resubmissions to other journals. Doing so is overall, a useful practice to efficiently use reviewer resources.

Best regards –
Andreas Gloeckner

Von: Andreas Gloeckner <andreas.gloeckner@uni-koeln.de>
Gesendet: Freitag, 25. November 2022 14:23
An: 'Gilad Feldman' <gfeldman@hku.hk>
Cc: 'Maximilian Maier' <maximilianmaier0401@gmail.com>
Betreff: AW: [Follow-up request to forward decision-letters to other journals]: Action Letter JDM[220823]-R1

Dear Dr. Feldman:

I asked the reviewers if they have objections against that and will answer as soon as I hear from them.

Best regards
Andreas Gloeckner

Von: Gilad Feldman <gfeldman@hku.hk>
Gesendet: Freitag, 25. November 2022 12:59
An: 'Andreas Gloeckner' <andreas.gloeckner@uni-koeln.de>
Cc: Maximilian Maier <maximilianmaier0401@gmail.com>
Betreff: [Follow-up request to forward decision-letters to other journals]: Action Letter JDM[220823]-R1

Dear Prof. Gloeckner,

Thank you for the reviews obtained and the decision letter.
We appreciate the feedback.

Follow up question/request:
In our submission email we explicitly asked that in case of a rejection we would be able to forward this decision to a different journal opting for a streamlined review (such as in Collabra:Psychology and Meta Psychology, who have open peer review). You did not comment on that in your decision letter.

May we proceed in forwarding your decision letter with our replies to those other journals?

Best regards,

--
Gilad Feldman (Fili)
Department of Psychology
University of Hong Kong

Website | Google Scholar | Mailing list & updates

From: Andreas Gloeckner <andreas.gloeckner@uni-koeln.de>
Sent: 24 November 2022 18:27
To: Gilad Feldman <gfeldman@hku.hk>
Cc: journal@sjdm.org
Subject: Action Letter JDM[220823]-R1

Dear Dr. Feldman,
dear reviewers in bcc:

I have now received feedback from two experts in the field concerning your revised manuscript "Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)" that you find attached below.

Both reviewers are experts in the field and identified major issues concerning the replication project and the manuscript in general. These are mainly in line with my previous concerns, which lead to an initial rejection of the manuscript. You aimed to address the concerns in a revision by some rewriting and decided against collecting new data. You could not convince the reviewers that your treatment of the issues was sufficient and they raise several further important limitations that I was not aware of. Both reviewers argue that this is not a reasonable replication given that the dependent variable changed considerably. They also make further important comments concerning problems, that are clearly formulated so that I do not have to reiterate them here.

Action: I reject the manuscript.

The paper might be alternatively presented as a test of the effect(s) in a different context / way (see R1). Also, the RoBMA-analysis included in the paper might be of value either in this or a separate manuscript, but the analysis also has substantial weaknesses (R1 & R2) and your descriptions are not sufficient to evaluate what was done (R1).

I appreciate your efforts to conduct replications of established effects in the field, but in this case the quality of the replication is too low to warrant publication in this journal. I am sorry that I cannot bring you better news on this manuscript.

Best regards –
Andreas Gloeckner

Reviewer 1

As I was asked to comment primarily on the re-analysis of the meta-analysis by Lee & Freely (2016) I focus on the respective part of the manuscript and discuss the replication itself only briefly.

In the introduction the authors argue convincingly that studies on the identifiable victim effect are in need of direct replications. The following study is meticulously and transparently documented, and the data analysis is very carefully done. However, it appears severely inconsistent to me to argue for a direct replication and then to run a study that changes the dependent variable. In the primary study the DV is "amount of money donated", in the replication it is "intended donations in a hypothetical scenario". Because of this change I am not sure whether the study presented in the paper can be considered a replication at all. I think that it would be perfectly reasonable to argue that both studies investigate two (hopefully) related but clearly different kinds of behavior ("donating money" vs "claiming to donate money"). It seems obvious to me that, in general, the factors affecting what people do with their money will be different from the factors affecting what people claim to do with their money. A priori, there is also very little reason to expect that a given factor will affect both kinds of behavior in the same way. Furthermore, the introduction calls into question the very existence of an identifiable victim effect. Given this state of affairs, a replication of studies investigating possible debiasing methods is rather not what is needed. This state also implies that extensions of studies investigating the identifiable victim effect are of secondary importance. This seems relevant to me as it suggests that the design of a meaningful and convincing replication study could be reduced to just two conditions. With only two conditions, however, it could have very well been possible to run a high-powered direct replication (with actual donations as DV) without any additional costs.
The paper is very clear about the change of the dependent variable between the original study and the replication. It also discusses problems that may be arising from this difference. This discussion, however, appears unspecific and not very helpful to me. After reading the paper, I still do not know why it should not matter whether or not participants can actually take some money home depending on their decisions in the study. I have also no idea why the paper does not present the study that it calls for in the introduction.

The re-analysis of the meta-analysis by Lee & Freely is based on RoBMA. This methodological decision is in line with the with the recent literature on methods for the detection and correction of publication biases and is well substantiated in the paper. I also think that there is good reason to believe that the main results of the re-analysis (i.e., that the meta-analysis is affected by publication bias and that this bias may inflate the effect size estimate in a relevant way) are correct. However, the report of this re-analysis is so very brief and incomplete that it is impossible to assess what actually happened (and for many readers the results are probably hard to understand). I simply list several pieces of information that are missing or that would be helpful for me. Large parts of this information could, of course, also be given in the supplementary materials – they should, however, not be available only in the form of sparsely commented code and a meta-analytical data set on the OSF.

- With regard to the underlying data, the number of included papers, studies, effect sizes and independent effect sizes should be given (if applicable: for each individual analysis). Furthermore, some information on the sample sizes of these studies would be helpful: I learned from the data set on the OSF that a study with a sample size of about N = 12.000 is included. This single study is (probably) larger than all the remaining studies taken together and, therefore, has to have tremendous influence on the results of the meta-analysis and its re-analysis. This should be transparent. Furthermore, I learned from the data set that the three largest studies all report effect sizes of (about) 0. This information alone suggests to me quite clearly that publication bias might be a problem.
- Given the change of the dependent variable in the following study, it would be very interesting to know if the meta-analysis also includes studies that use "donations in a hypothetical scenario" as DV. If so, then the obvious question is whether the difference between actual and hypothetical donations is a moderator. If no such studies are included that should be stated very clearly. – Again, in this case I am not sure of what relevance the meta-analysis and its re-analysis are for the remainder of the paper.
- With regard to the RoBMA-analysis itself, the main problem is that it remains unclear which data models are included in the averaging procedure of RoBMA. As the results of RoBMA have to depend on the considered data models this information is essential. Furthermore, the included models need to specify prior distributions for their parameters. These prior distributions should be given (alternatively: it should be stated from which source these prior distributions are taken).
- Footnote 2 refers to an additional analysis that uses only "the most precise estimate within each experiment". From the meta-analytical data set it appears, however, that most effect size estimates from the same experiment are based on the same sample size. Thus, in these cases, I have no idea which estimate is more precise and included in the analysis.

- The estimates of RoBMa depend on the weightings of different models which in turn depend on the fit between these models and the data. Therefore, to fully understand the RoBMA results it would be very helpful to have some indication of these weightings (which could be given in a table).
- Figure 2 does not appear very helpful to me. The left panel is misleading to some degree as it shows the effect size estimate of PET-PEESE which is not the RoBMA estimate that is actually interpreted. As the weightings of the different models are not known it is unclear which impact the PET-PEESE estimate has on the RoBMA estimate (i.e., the figure may show the estimate of a model with a very poor fit). – I do not really understand the right panel, which is at least partly due to the missing information about the included models. Does the constant publication probability with p >.5 imply that only one-sided selection models were considered? If also two-sided models were included why does the estimated publication probability remain constant at high thresholds (p = .95, p =.975)?

Frank Renkewitz
(signed review)

Reviewer 2

Let me begin with what I like about the paper. Replication is generally important and useful and clearly there are many methodological improvements in the current paper—larger sample sizes, pre-registration, open data, etc that were not standard when Small et al. (2007) was conducted and published.

That said, the fact that this replication involves hypothetical donation undermines the objectives of replication. In my view, a hypothetical donation is a poor substitute for a few reasons. First, participants are less likely to take the task seriously when there is no money at stake. This is generally true for hypothetical choice and a key reason why experimental economics as a field insists on incentive-compatible designs (eg., Levitt and List, 2007). Second, when studying generosity specifically it seems especially critical. Generosity is about self-sacrifice but a hypothetical donation does not involve any sacrifice. That is, it is easy (cost-less) to reply in a survey that you would donate a lot. Indeed, as the authors note, participants in these replication studies report that they would hypothetically donate much more than what is found in typical donation studies, including in Small et al. (2007). Typically, in dictator game studies and real donation studies there are a high proportion of 0s and often a higher proportion of the max (5 in this case) when the scale is low and fixed as it was in this case.

Another difference (not discussed in the replication paper) is that in Small et al., the study was designed with a cover story that the donation request was not part of a 'study'. That is participants were asked to complete an unrelated study on technology usage and then paid for that study. Upon payment, they were given a letter from a charity, an envelope and 5 $1 bills. In contrast, in the replication study, participants can only infer that their charitableness is what the researchers are examining.

The authors discuss the hypotheticality as a limitation, so the question is how severe is this as a limitation and what is the resulting contribution of the paper. My view is that this is not a face valid way of assessing generosity. No paper is perfect and perhaps this could be chalked up as a 'limitation' if there were some greater theoretical or practical contribution of the paper. But the paper's main goal is to examine whether this effect replicates, and that is tested in a way that although methodologically superior to the original studies in certain ways (pre-registration, larger samples, etc,) is critically inferior in this other way.

Other comments:

1. From abstract: "These findings suggest that the identifiable victim may be

better framed in terms of 'scope-insensitivity'. This doesn't seem the right way to describe this particular effect. Scope insensitivity would be if there was one child or many that are supported, such that there is evidently more social utility to helping many (e.g., Kogut and Ritov studies where there is medicine for 1 or 10 children). In the current study, the 'statistical victim" condition merely describes statistics of the magnitude of need. It does not imply that all of the people in the statistics will be supported in the same way that the child in the identifiable victim condition will be.

This also relates to the discussion at the end of the manuscript about whether it reflects irrational decision making. Surely, using some pool of resources to benefit one person that could alternatively be used to provide the same benefit to >1 would reflect a failure of utilitarian reasoning and thus "irrational" in some sense. But I don't think you can evaluate that question given the current operationalization of the IVE.

2. Please note also that the Lee and Feely meta-analysis, which is modified and argued to provide further evidence of a null effet, did not include large-scale field studies that find evidence consistent with the IVE (Galak et al, 2011; Sudhir et al., 2016)

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance Decision Making: A Field Study of Prosocial Lending. Journal of Marketing Research, 48(SPL), S130–S137. https://doi.org/10.1509/jmkr.48.SPL.S130

K. Sudhir, Roy, Cherian (2016). "Do Sympathy Biases Induce Charitable Giving? The Effects of Advertising Content" , Marketing Science, https://doi.org/10.1287/mksc.2016.0989

3. I found Table 1 overview of hypotheses to be confusing:

? What does "main effect" in H1 mean? Shouldn't this hypothesis about a simple effect absent any intervention?

? What drives H2? Why would one expect a main effect of learning? The original paper hypothesized an interaction effect, more akin to H3.

? Why is H4 described as a main effect?

? Hb is confusingly worded. I'm having trouble interpreting it.

Prof. Dr. Andreas Glöckner (home)
Chair of Social Psychology, University of Cologne
Director of the Department Psychology (department head)
Richard-Strauss-Str. 2 (Room 2.A11), D-50931 Cologne, Germany, phone +49-221 470 7916

Senior Research Fellow, Max Planck Institute for Research on Collective Goods
research profile: https://scholar.google.de/citations?user=n5dYy5sAAAAJ
editor: Judgment and Decision Making

Reply to JDM decision letter reviews: Small et al. (2007) replication and extension

We would like to thank the JDM editor and the reviewers for their useful suggestions and their willingness to forward their decision letter to a different journal.

Below we provide a detailed response to each of the items. 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/phiAqhwGwKEu and the file "combined_comparison.pdf"

Response to Editor: Dr./Prof. Andreas Gloeckner initial desk review

[We resubmitted our manuscript with the following replies to the initial rejection, and the editor decided to proceed to sending it to review, which received a second rejection, attached below]

> **Thank you for submitting your manuscript "Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)" to Judgment and Decision Making.**
>
> **I have read the manuscript with interest. I found the meta-analysis interesting and also the citations of previous failed replications of the effect, although the reporting of the latter seems a bit unbalanced in the paper. I, however, see a fundamental problem in the methodology of the empirical replication:**
> **You basically replicate a real-effort real donation study with a pure hypothetical study and don't find any / much effects. Given this fundamental difference in the design, you conclusions go way to far. You may conclude that the identified victim effect does not replicate in hypothetical decision (with MTURK persons that might only care for their payment).**

Thank you, we understand and accept. This is an important comment.

We changed the language throughout the manuscript to emphasize that we are focusing on hypothetical decisions, in addition we changed the classification of the replication to a 'close to far' replication. To our knowledge and having consulted in experts in this domain, there is a little research on how hypothetical vs real money decisions affect donation judgements in practice; however, we added a short discussion based on the research available tying hypothetical donations to real-life findings.

We also greatly toned down our discussion and conclusions accordingly.

> **Also you manipulation check questions – which seem to be presented in between the materials - do not seem entirely innocent since they might disrupt feelings of empathy etc. But this is probably less of a problem.**

Given our experience in such tasks and this sample, we felt this as a needed check. We have rarely seen this affect DVs in other settings, other than decreasing noise that has to do with inattentiveness. We have previously noted this as a deviation yet welcome the request to do better in discussing this point. We now mention this in the discussion section.

> **Also your results concerning publication bias did not convince me on first glance, although I would need to ask an expert on that. I have followed the usage of RoBMA as an interesting approach in other publications. In this case, however, inspection of Figure 1 seems to indicate that the regression line is driven mainly by three outliers (in both directions), whereas the other studies seem relatively well behaved. It would be quite concerning for various of the recent publication with RoBMA if this promising method does not include any sensitivity / outlier influence analysis.**

Thank you for this observation. From our read of the funnel plot it appears that this refers to outliers in terms of precision/sample size. We do not believe that the outliers should be excluded in this case as the include the two highest powered studies. Dropping the two largest studies from the meta-analysis would result in losing information. Typically, when excluding outliers in PET-PEESE these would be outliers in terms of effect size rather than precision.

However, we took this to heart and did our best to address this concern by adding an analysis that only relies on the *p*-value models (selection models) of RoBMA (i.e., not dependent on any regression of effect sizes on standard errors at all). This analysis seems to corroborate the findings of the main analysis.

Response to Editor: Dr./Prof. Andreas Gloeckner final decision letter review

> **Both reviewers are experts in the field and identified major issues concerning the replication project and the manuscript in general. These are mainly in line with my previous concerns, which lead to an initial rejection of the manuscript. You aimed to address the concerns in a revision by some rewriting and decided against collecting new data. You could not convince the reviewers that your treatment of the issues was sufficient and they raise several further important limitations that I was not aware of. Both reviewers argue that this is not a reasonable replication given that the dependent variable changed considerably. They also make further important comments concerning problems, that are clearly formulated so that I do not have to reiterate them here.**
>
> **Action: I reject the manuscript.**

Thank you for the reviews obtained and your feedback.

We believe that the concerns raised are not about the methods or the rigor of the paper, but rather about diverging views of what to expect from a replication and meta reanalysis project of this phenomenon. We believe that we have been very careful in the way we addressed the issues, yet welcome the opportunity to further clarify and elaborate on our decisions.

We strongly believe that the academic community would greatly benefits from the empirical evidence communicated in this manuscript, as we move forward in the investigation of the identifiable victim effect, and hope that other journals would share that view.

> **The paper might be alternatively presented as a test of the effect(s) in a different context / way (see R1). Also, the RoBMA-analysis included in the paper might be of value either in this or a separate manuscript, but the analysis also has substantial weaknesses (R1 & R2) and your descriptions are not sufficient to evaluate what was done (R1).**
> **I appreciate your efforts to conduct replications of established effects in the field, but in this case the quality of the replication is too low to warrant publication in this journal. I am sorry that I cannot bring you better news on this manuscript.**

Below we address each of the issues raised in detail. We believe that we were able to address these concerns. We do not think that the categorization of the replication as being low quality is fair, as there were no questions raised regarding the methods or rigor, but rather the reviews were

focused on the diverging views regarding what is expected from this replication or how to best interpret our findings.

In addition, we have been very clear on everything that we conduced, and shared all materials, data, and code that allow anyone interested in examining all that we did meeting and exceeding what is currently typical in our field. We welcome opportunities to improve and further addressed raised points regarding further need to clarify our sharing and disclosures, yet we disagree that it was not possible to evaluate what we conducted, we believe we have met very high standards of transparency. As evidence for that we previously noted and provided an external 3rd party evaluator report that validated all of our analyses who was provided the same materials as those submitted with our manuscript.

We further elaborated on our decision to study intent and hypothetical scenarios, and given the feedback changed our categorization of the replication to “far” and “conceptual”.

Response to Reviewer #1: Dr./Prof. Frank Renkewitz

R1.1) As I was asked to comment primarily on the re-analysis of the meta-analysis by Lee & Freely (2016) I focus on the respective part of the manuscript and discuss the replication itself only briefly.

Thank you very much for the detailed and constructive review. Much appreciated!

R1.2) In the introduction the authors argue convincingly that studies on the identifiable victim effect are in need of direct replications. The following study is meticulously and transparently documented, and the data analysis is very carefully done. However, it appears severely inconsistent to me to argue for a direct replication and then to run a study that changes the dependent variable. In the primary study the DV is "amount of money donated", in the replication it is "intended donations in a hypothetical scenario". Because of this change I am not sure whether the study presented in the paper can be considered a replication at all. I think that it would be perfectly reasonable to argue that both studies investigate two (hopefully) related but clearly different kinds of behavior ("donating money" vs "claiming to donate money"). It seems obvious to me that, in general, the factors affecting what people do with their money will be different from the factors affecting what people claim to do with their money. A priori, there is also very little reason to expect that a given factor will affect both kinds of behavior in the same way. Furthermore, the introduction calls into question the very existence of an identifiable victim effect. Given this state of affairs, a replication of studies investigating possible debiasing methods is rather not what is needed. This state also implies that extensions of studies investigating the identifiable victim effect are of secondary importance. This seems relevant to me as it suggests that the design of a meaningful and convincing replication study could be reduced to just two conditions. With only two conditions, however, it could have very well been possible to run a high-powered direct replication (with actual donations as DV) without any additional costs. The paper is very clear about the change of the dependent variable between the original study and the replication. It also discusses problems that may be arising from this difference. This discussion, however, appears unspecific and not very helpful to me. After reading the paper, I still do not know why it should not matter whether or not participants can actually take some money home depending on their decisions in the study. I have also no idea

> **why the paper does not present the study that it calls for in the introduction.**

Thank you.

There are a lot of points made here and we will try and address each:

Goals of the replication:

We indeed changed the dependent variable and have been very transparent and clear about that and therefore classified this as being between close and far replication. Aside from the change in dependent variable we were careful to try and follow the rest of the original's vignettes. We may have not been clear enough about why we made that change, and why we expected the effect to hold, and we are grateful for the encouragement to elaborate on that. We added the following:

> "We note that we initially set out to conduct a direct close replication, yet decided on first running a far conceptual replication using the same design yet with an important adjustment of the dependent variable to use hypothetical donations, which can be approximated as intent, rather than real donations. We did this for a number of reasons. First, this project followed on a different replication project we conducted in Majumder et al., (2023) of the identifiable victim effect by Kogut and Ritov (2005) than showed the effect using hypothetical donations, as many other studies examining the identifiable victim effect have. We therefore aimed to make the two replications as close as possible in their dependent variable. Second, we acknowledge the difference between intent and real-life behavior, yet thought it best to ensure the effect holds with simpler intent before embarking on the more complex and costly behavior, given that effects for donations typically seem stronger for intent than for behavior and seem higher up in the causal chain. Given this important adjustment regarding the dependent variable, though much of the rest of the study remains the same, we categorized this replication as far and conceptual. We caution against over-interpreting from this replication to the likelihood of original article's real donations effect replicating, though we hope the community would find this informative in the generalizability of the original's design to hypothetical scenarios and intent. We discuss this point and implications in the general discussion."

You argued that given the mixed findings in the literature was an exact very close replication of the target article might be warranted, and we indeed call for such a replication in our general discussion. Yet, the question here is whether this must be the goal of this replication, and we do not believe that not meeting the categorization of an very close replication means that the evidence we provide is of lesser value, as is too is informative regarding the generalizability of the phenomenon. We were hoping to increase chances for a successful replication, as we thought real stakes behavior with no underlying low-stakes intent unlikely, with intent more likely to show the effect. Though not unheard of, it would seem strange that an effect would show for

behavior and not for intent, especially given that much of that literature, like Kogut and Ritov (2005) has been shown on hypothetical scenarios and intent.

Classification of the replication:

Given the change in the DV, in our manuscript, we used the Lebel et al. (2018) categorization to classify our replication as being between "close and far" given that the methods closely resemble the original's yet the change in the DV is one that makes interpretation more difficult.

In our revised submission we went further to remove mention of direct close replications, and classify our replication attempt as conceptual.

Alignment with literature's needs

Both direct and conceptual replications are needed, and we thought that our replication combined the best of these worlds. We think that a follow-up very close direct replication is needed, which we called for in our discussion, and our results now help adjust our priors regarding the likelihood of that succeeding so that we may aim to detect smaller effects and think of ways in which we can help ensure a direct replication would be more likely to succeed. Our replication is a needed step in informing the closer real-life replication. Our findings also suggest that it would be valuable for future research to further investigate the donation intent-behavior link.

> **The re-analysis of the meta-analysis by Lee & Freely is based on RoBMA. This methodological decision is in line with the with the recent literature on methods for the detection and correction of publication biases and is well substantiated in the paper. I also think that there is good reason to believe that the main results of the re-analysis (i.e., that the meta-analysis is affected by publication bias and that this bias may inflate the effect size estimate in a relevant way) are correct.**

Thank you, we are glad to receive a second confirmation that our analyses are relevant and accruate.

> **However, the report of this re-analysis is so very brief and incomplete that it is impossible to assess what actually happened (and for many readers the results are probably hard to understand). I simply list several pieces of information that are missing or that would be helpful for me. Large parts of this information could, of course, also be given in the supplementary materials – they should, however, not be available only in the form of sparsely commented code and a meta-analytical data set on the OSF.**
>
> **R1.3) With regard to the underlying data, the number of included papers, studies, effect sizes and independent effect sizes should be given (if applicable: for each individual analysis). Furthermore, some information**

> **on the sample sizes of these studies would be helpful: I learned from the data set on the OSF that a study with a sample size of about N = 12.000 is included. This single study is (probably) larger than all the remaining studies taken together and, therefore, has to have tremendous influence on the results of the meta-analysis and its re-analysis. This should be transparent. Furthermore, I learned from the data set that the three largest studies all report effect sizes of (about) 0. This information alone suggests to me quite clearly that publication bias might be a problem.**

Thank you. Everything we did is transparent, as others – like you – could see all that we did on the OSF. We appreciate the nudge for further clarifications in the manuscript.

We added information about the number of experiments and effects included and the information about the high powered study as well as that the three largest studies show mostly null effects.

> "Lee and Freely (2016) conducted a meta-analysis that summarized 41 effects from 22 experiments on the identifiable victim effect. They conclude that across these studies, they find a "significant yet modest IVE [identifiable victim effect]" (Lee & Freely, 2016, p. 199). However, there is indication that this effect might be much weaker if publication bias is accounted for. The three largest powered studies in the dataset show effect that are almost zero, including one study with 12 802 participants (r = 0.004)."

> **R1.4) Given the change of the dependent variable in the following study, it would be very interesting to know if the meta-analysis also includes studies that use "donations in a hypothetical scenario" as DV. If so, then the obvious question is whether the difference between actual and hypothetical donations is a moderator. If no such studies are included that should be stated very clearly. – Again, in this case I am not sure of what relevance the meta-analysis and its re-analysis are for the remainder of the paper.**

This is a good point, and one that we are happy to elaborate on. Since we submitted both of our identifiable victim replications to JDM, the editor was aware of our other project, but we should have been clearer about our goals and why we made the adjustment to the dependent variable. Indeed, some of the literature, like Kogut and Ritov (2005) which we attempted to replicate in our other project, used hypothetical donations rather than real donations, and we aimed to bring the two closer together to try and get at why our replication of Kogut and Ritov may have failed.

Unfortunately, the meta-analysis we re-analyzed did does not look at moderation regarding hypothetical versus real donations, though it did include both types of studies. (For a study to be included, it had to meet the following inclusion criteria: […] (5) employ a measure of intention to contribute money or time, or a measure of actual contribution of money as the dependent variable, Lee and Freely, p. 202). The meta-analysis is not the focus of our investigation here,

and so supplementary coding and moderation is beyond the scope we intended. We agree that this is a valuable avenue for future research.

> **R1.5) With regard to the RoBMA-analysis itself, the main problem is that it remains unclear which data models are included in the averaging procedure of RoBMA. As the results of RoBMA have to depend on the considered data models this information is essential. Furthermore, the included models need to specify prior distributions for their parameters. These prior distributions should be given (alternatively: it should be stated from which source these prior distributions are taken).**

We used the default version of RoBMA as in all other RoBMA applications. As it includes 36 models, and this is not the main focus on our manuscript, we did not consider it useful to explain each of these models in detail. Instead we referred to the paper, which introduces this version of RoBMA:

> "Here we use the version of RoBMA (also known as RoBMA-PSMA [Publication Selection Model Averaging]) as in Bartoš, Maier, Wagenmakers et al., (2022) as it has been vetted extensively in simulation studies and applied examples (in the same paper). For details about the 36 models that are included, and the corresponding prior distributions and prior model probabilities please see Bartoš, Maier, Wagenmakers et al., (2022)."

> **R1.6) Footnote 2 refers to an additional analysis that uses only "the most precise estimate within each experiment". From the meta-analytical data set it appears, however, that most effect size estimates from the same experiment are based on the same sample size. Thus, in these cases, I have no idea which estimate is more precise and included in the analysis.**

Thank you. We updated this to bootstrap across the effect sizes with the same precision, it is now explained in the footnote 2:

> "Due to the lack of publication bias correction methods that can accommodate a three-level structure, we accounted for the dependency by only using the most precise estimate within each experiment. Often there were multiple estimates with the same precision within a study. In this case, we selected randomly and bootstrapped 500 times. Using the median of these bootstraps, this analysis comes to the conclusions regarding evidence for publication bias and evidence for an effect. Unlike the main analysis we find moderate rather than weak evidence against heterogeneity."

> **R1.7) The estimates of RoBMa depend on the weightings of different models which in turn depend on the fit between these models and the data. Therefore, to fully understand the RoBMA results it would be very helpful to have some indication of these weightings (which could be given in a table).**

RoBMA includes 36 models. Therefore, we thought it out of scope to include all of them in the main text. However, we added a table with Bayes factors and the posterior probabilities for the 36 models in the supplement, and those are provided in our provided code.

> **R1.8) Figure 2 does not appear very helpful to me. The left panel is misleading to some degree as it shows the effect size estimate of PET-PEESE which is not the RoBMA estimate that is actually interpreted. As the weightings of the different models are not known it is unclear which impact the PET-PEESE estimate has on the RoBMA estimate (i.e., the figure may show the estimate of a model with a very poor fit). – I do not really understand the right panel, which is at least partly due to the missing information about the included models. Does the constant publication probability with p >.5 imply that only one-sided selection models were considered? If also two-sided models were included why does the estimated publication probability remain constant at high thresholds (p = .95, p =.975)?**

Figure 2 is the standard figure as included with all RoBMA publications. Therefore, we kept it. The estimate from PET-PEESE contributes to the model-ensemble, therefore, we believed it would be useful to keep it. We now give the posterior probability for all selection models vs PET-PEESE models. We explain that the right panel shows the model-averaged selection probabilities across the different selection models included in RoBMA.

The right panel displays estimates for the relative publication probabilities of nonsignificant studies in comparison to significant studies model averaged across the different selection models included in RoBMA

Response to Reviewer #2: Anonymous

> **R2.1) Let me begin with what I like about the paper. Replication is generally important and useful and clearly there are many methodological improvements in the current paper—larger sample sizes, pre-registration, open data, etc that were not standard when Small et al. (2007) was conducted and published.**

Thank you for the positive and constructive review, and the encouraging opening note.

> **R2.2) That said, the fact that this replication involves hypothetical donation undermines the objectives of replication. In my view, a hypothetical donation is a poor substitute for a few reasons. First, participants are less likely to take the task seriously when there is no money at stake. This is generally true for hypothetical choice and a key reason why experimental economics as a field insists on incentive-compatible designs (eg., Levitt and List, 2007). Second, when studying generosity specifically it seems especially critical. Generosity is about self-sacrifice but a hypothetical donation does not involve any sacrifice. That is, it is easy (cost-less) to reply in a survey that you would donate a lot. Indeed, as the authors note, participants in these replication studies report that they would hypothetically donate much more than what is found in typical donation studies, including in Small et al. (2007). Typically, in dictator game studies and real donation studies there are a high proportion of 0s and often a higher proportion of the max (5 in this case) when the scale is low and fixed as it was in this case.**

We understand. Please see our response to R1.2 above.

In addition, this seems to be a broader critic of the JDM literature and the intent based identifiable victim effect studies. While behavioral economics is focused on real incentives, the Judgment and Decision Making literature is not, and the psychological science literature cares about intent. In addition, much of the behavioral economics literature originated and is based on initial demonstration using very simplified vignette studies.

Finally, what you wrote was one of the reasons why we opted for intent before proceeding to behavior. Behaviors are more difficult to observe, especially regarding self-sacrifice donations, and so real-behavior is more conservative and less likely to observe as it involved many factors that go beyond decision-making.

> **R2.3) Another difference (not discussed in the replication paper) is that in Small et al., the study was designed with a cover story that the donation request was not part of a 'study'. That is participants were asked to complete an unrelated study on technology usage and then paid for that study. Upon payment, they were given a letter from a charity, an envelope and 5 $1 bills. In contrast, in the replication study, participants can only infer that their charitableness is what the researchers are examining.**

Again, this makes the real-behavior more conservative and less likely to be observed. If the participants could infer what we are examining, and would have liked to aligned with that, we would have been more rather than less likely to observe the effect.

About the specific argument about the separate study. We think it is quite unlikely that participants did not realize that their charitable donations were investigated. Instead, the real effort task is often used to calibrate participants subjective valuation of money in terms of the time they invested to earn it.

To address the need to note this more clearly, we now mention this in the limitations section.

> "Second, given this cover story participants may not have realized that the experimenters were investigating their donation behavior."

> **R2.3 The authors discuss the hypotheticality as a limitation, so the question is how severe is this as a limitation and what is the resulting contribution of the paper. My view is that this is not a face valid way of assessing generosity. No paper is perfect and perhaps this could be chalked up as a 'limitation' if there were some greater theoretical or practical contribution of the paper. But the paper's main goal is to examine whether this effect replicates, and that is tested in a way that although methodologically superior to the original studies in certain ways (pre-registration, larger samples, etc,) is critically inferior in this other way.**

See our responses to the previous comments.

Other comments:

R2.4) From abstract: "These findings suggest that the identifiable victim may be better framed in terms of 'scope-insensitivity'. This doesn't seem the right way to describe this particular effect. Scope insensitivity would be if there was one child or many that are supported, such that there is evidently more social utility to helping many (e.g., Kogut and Ritov studies where there is medicine for 1 or 10 children). In the current study, the 'statistical victim" condition merely describes statistics of the magnitude of need. It does not imply that all of the people in the statistics will be supported in the same way that the child in the identifiable victim condition will be.

This also relates to the discussion at the end of the manuscript about whether it reflects irrational decision making. Surely, using some pool of resources to benefit one person that could alternatively be used to provide the same benefit to >1 would reflect a failure of utilitarian reasoning and thus "irrational" in some sense. But I don't think you can evaluate that question given the current operationalization of the IVE.

Think about it.

R2.5) Please note also that the Lee and Feely meta-analysis, which is modified and argued to provide further evidence of a null effet, did not include large-scale field studies that find evidence consistent with the IVE (Galak et al, 2011; Sudhir et al., 2016)

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance Decision Making: A Field Study of Prosocial Lending. Journal of Marketing Research, 48(SPL), S130–S137. https://doi.org/10.1509/jmkr.48.SPL.S130

K. Sudhir, Roy, Cherian (2016). "Do Sympathy Biases Induce Charitable Giving? The Effects of Advertising Content" , Marketing Science, https://doi.org/10.1287/mksc.2016.0989

Thank you for providing these references.

We chose not to update the meta-analysis as in this case it would be difficult to justify why we updated it with these studies and not with others. However, me now mention these field studies as successful replication of the identifiable victim effect.

R2.6) I found Table 1 overview of hypotheses to be confusing: What does "main effect" in H1 mean? Shouldn't this hypothesis about a simple effect absent any intervention?

This is a main effect as it is collapsing across the intervention, we chose to test this effect rather than the simple effect as we can test it with more power.

Why is H4 described as a main effect?

H4 is with regards to the joint condition. Therefore, this is a main effect comparing identifiable victim condition to joint condition.

H5b is confusingly worded. I'm having trouble interpreting it.

We changed it to:

> "Learning about the identifiable victim effect reduces affective feelings, and the effect is stronger for an identifiable victim and joint presentation of identifiable victim with victim statistics, than for statistical victims"

Peer Review and Communication History
Streamlines Review

MS Title: Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)

Author Names: Maximilian Maier, Yik Chun (Nicole) Wong, Gilad Feldman

Submitted: Dec 24, 2022

Editor First Decision: Revise & Resubmit
Feb 19, 2023

Dear Dr. Feldman,

I apologize for the delay in getting this decision to you. Although your manuscript was streamlined, and did not go out for external review, I needed to find time to read everything carefully. From past experience with streamlined manuscripts (many from your lab!), I know that this often takes me an entire workday, and sometimes it takes a while before I'm able to clear a large chunk of time from my calendar. I know this probably does not alleviate any frustration you might have about the delay, but I will say that digging into a streamlined manuscript is one of my favorite ways to spend my time, so I look forward to the opportunity to spend an entire day immersed in that activity. This was no exception – your manuscript did not disappoint. It is a very rich submission, with many strong aspects, and overall I feel that it is very close to publishable. Nevertheless, there are some revisions that I am asking for before accepting your manuscript. I hope you will agree that they will improve the manuscript, and I believe they are all fairly straightforward to implement. Thus, I am happy to commit to accepting your manuscript if you are able to address these points, and if I find no or few new problems/errors in your revised submission.

I've numbered my points below, but many of them are interrelated and there is some repitition. I apologize for this - I wrote this decision letter as I was reading the manuscript, and tried to go back and make it more coherent, but there is still some redundancy, and the order of the points is relatively meaningless (they are not ordered from most to least important).

1. In some places you refer to your outcome measure for donations as measuring "intentions" or "intended donations", but I don't think that's accurate. From my reading of the DV, these are not intentions to donate, just hypothetical donations. That undermines the strength of some of your

arguments about the far replication still speaking to the effect in the original. To be clear, I agree that these results, with this hypothetical donation DV, are still relevant and worth publishing, and that the community should care about them and their implications for the original effect. And since you describe the study as a far replication, there isn't much change needed in the framing. However, I think you do need to change your language where you refer to intentions or a similar concept, and you need to tone down some of the arguments you make that depend on this conceptualization of your DV. After having read the reviews and response to the reviews from JDM, I want to add that I think there is probably an interesting discussion to be had about what the relationship is between answers to hypothetical scenarios, and intentions. I'd be open to you making an argument that there is a connection, but at least to me, as someone outside of JDM who interprets "intent" in the common sense way, my answer to a hypothetical scenario is not an intention, and I think that perspective should be represented in your framing.

2. It wasn't clear to me from the study description what participants in the implicit learning/joint statistical and identifiable information condition were asked to do. Were they only asked to donate to the identifiable victim? Were they given a choice about which victim(s) they could donate to? I found the answer to this question (only asked to donate to the identifiable victim) on page 19, but this information is necessary to interpret the research questions and hypotheses, so it needs to be explained in greater detail when the study design is described (pp. 12-15), and perhaps even when describing the results of Small et al. (pp. 11-12), because it's hard to make sense of this condition without knowing what the donation target was.
3. Related to this, I think it would help a lot if, when describing the Small et al. hypotheses and findings (pp. 11-12), you explicitly describe the design of each of the two studies. I had to go look these up to understand how your 2 x 3 design maps onto their designs, and I believe their designs are: Study 1 is a 2 (identifiability: identifiable vs. statistical) x 2 (explicit learning: intervention vs. control) between-subjects factorial design, and Study 3 has only one factor with three levels/conditions: identifiable victim, statistical victims, and the joint/implicit learning condition where the donation target is an identifiable victim but statistical information is also provided.
4. One of my biggest struggles throughout the paper was figuring out which cells in your design were compared to test each hypothesis. In Table 1, you describe the hypotheses verbally, but it wasn't clear to me how this mapped onto the different cells in your design. What would help is if you would move Table 3 earlier, remove the text about hypothetical donations and affective feelings that is currently in the middle of the table (this text is misleading anyway, as it seems like the DVs are paired with the rows of IV2, when in fact

all DVs were measured in all conditions), and instead fill in the table with a label for each of the six cells (e.g., "A1, A2, A3" in the top row and "B1, B2, B3" in the bottom row, or even just the letters A through F, or whatever you think is best). Then. in the table describing your hypotheses (what is currently Table 1), refer to these cells explicitly (in addition to the verbal description of the hypothesis). For main effects, this might be something like "comparing cells A1 and A2 vs. cells B1 and B2", and for an interaction this might be something like "the difference between cell A1 and B1 will be greater than the difference between cells A2 and B2".

5. Related to this, you refer to the two factors as "identifiability" and "intervention" but I think this isn't quite correct. One factor is just whether there was the explicit intervention (explicit learning vs. control). But the other factor is a combination of the simple identifiable vs. statistical identifiability variable, and the implicit learning intervention, which happens to be carried out by combining the identifiable and statistical information, but is conceptually an intervention, not a level of identifiability. This makes your design quite messy, because the two factors of the design don't map on neatly to two conceptual (causal) variables. I'm not sure how best to label the two variables given this situation. I am open to you keeping the oversimplified "intervention" and "identifiability labels", but it would help if you explicitly acknowledged that one of the identifiability levels is the implicit learning intervention in Small et al.'s Study 3 (I think). In addition, I suggest changing the "intervention" label to "explicit learning intervention" or something like that.
6. In the results section, you say that you followed (and extended) the analyses conducted by the target article, but because your design was different (2 x 3, which meant that you ended up crossing the explicit learning conditions with the implicit learning conditions, which I believe Small et al. never did, since those two manipulations were in separate experiments?), it is not clear what the most similar analyses would be. This becomes a bit clearer when you report the analyses for each hypothesis (starting on page 23), but even then it is not totally clear. For example, on page 23 you describe a 2 x 2, and I can guess which of the six cells got dropped (the two cells in the joint level of identifiability), but you never say this explicitly (and indeed, your Figure 2 includes all of the cells, unlike the analysis, which is only explained indirectly - it needs to be more explicit). If you alert readers early on (in the intro around pp. 12-15, where I am asking you to add a table with your design and labels for each cell) that your design makes it tricky to map each cell onto the similar/identical cell in the original studies, and then remind readers at the beginning of the results section of this challenge, and of the fact that you will need to compare subsets of cells (not the whole 2x3 design) for many of your analyses, I think this will help a lot. Then, when you get to each specific

analysis (p. 23 onwards), you can once again use the labels I've asked you to create for each cell to make it super clear to readers which cells are being compared or dropped in each analysis.
Related to this, I could not make sense of Table 8. I admit that if I tried harder, I probably could get there, but I don't think many readers are going to try harder than I have. I think adding the labels for each cell, and then explicitly saying which cells are being compared to which for each analysis, will help a lot with Table 8.

7. Separate from the issue of being clearer about which cells are included in which analysis, I am concerned that your analyses, if I understand them correctly, did not always use the "right" cells, if the goal is to match the analyses in Small et al.For example, for the bottom half of the first row titled "Identifiability (with joint condition) [Study 3]", it doesn't sound to me like these analyses match the Small et al. Study 3 analyses, because Study 3 did not include the explicit learning intervention, right? From the text on page 24, it sounds like this analysis includes all six cells, rather than only including the three cells in the control level of the intervention variable, which is what you'd need to do to match the Small et al. Study 3 test of a main effect of identifiability.From looking at the descriptive results, I believe that when you limit the test of the effect of identifiability to include only the cells in the control condition of the intervention variable, you will find even smaller effects/stronger evidence for the null. This is important - to the extent that there is any effect of the joint condition, it seems to be driven by the cells in the intervention condition, which means that presenting the joint information only had an effect (to the extent that it had an effect at all, which is very small) when presented together with the explicit learning information. I realize the interaction was not significant, but my main point is that I think the appropriate test of the effect of implicit learning/joint presentation is a simple 3-cell comparison of the control condition cells, and this will likely show strong evidence for the null.I suspect that a similar decision (about how to alter the analyses to best match the original analyses and the conceptual question) will come up for other hypotheses, but I haven't considered/thought through each of them carefully, in part because the current structure of the results section makes it hard for me to think through what each hypothesis is and what the best test would be (see my next point). These issues also apply to the analyses of the feelings aggregate, but here the situation is more complicated because, if I understand correctly, feelings were only measured in Study 1 in Small et al., so the most appropriate test if we want to compare the results to the original would be excluding the two cells in the joint level of the identifiability variable. However, it might be interesting to readers to know how the joint condition affected feelings, so it might also be worth doing the three-cell comparison of

feelings across the three cells in the control level of the intervention variable, which would also have the advantage of matching the analyses you'd be conducting for the donation DV. On the other hand, the descriptive results in Table 7 are enough for curious readers to look at this if they want to (especially since your data will be publicly available), so perhaps you could just stick to the 2x2 that matches the analysis of feelings in Small et al., Study 1.If changing some of the analyses to better match the original studies leads to a deviation from your preregistered analyses, you will need to decide whether/where to present the preregistered analyses that I am asking you to replace. My strong preference would be that they be moved out of the main text, so that the main text only includes the most appropriate test of each hypothesis (i.e., the test that most closely matches the original), even if these were not preregistered. But it should also be very clear to readers, without having to read the supplement, what was and what was not preregistered.My reason for requesting that you change the analyses to better match the original analyses is not just out of a desire to match the original studies. I also think these tests are the right ones to test the conceptual hypotheses. That is, for example, if I want to know if the effect of the joint condition, I think the implied research question here is, "in the absence of any other interventions, does presenting the statistical information in addition to the identifiable information impact donations?". Thus, only the cells in the "control" level of the (explicit) intervention condition are relevant. This is what I assume most readers would be interested in, and what Small et al.'s Study 3 tested. Moreover, it's confusing to me that for the main effect of identifiability, you did drop some cells to make the analysis match Small et al.'s Study 1 (i.e., you dropped the two joint condition cells), but for the effect of identifiability, you didn't drop the cells that should be dropped to match Small et al.'s Study 3 (i.e., the explicit intervention condition cells). The fact that you are inconsistent about this is a further reason to revise your approach, and make it more consistent throughout, and consistent with your stated aim of matching the analyses of the original paper.

8. Another issue is that the results section lacks a clear structure. For example, the heading on page 23 is about an interaction effect, but the first paragraph is about a main effect (mostly, it also talks about some things that are not specific to the main effect, such as the figure and the Bayesian analytic approach). I would like you to structure the results section to match the structure of the hypotheses presented in Table 1, with a heading or subheading for each hypothesis, and text under each heading that matches the heading. I realize that some things you want to report in the results section are not specific to one hypothesis (e.g., some figures, and the description of the analytic approach, etc.) - these can be presented at the

beginning, before the results speaking to each specific hypothesis, or wherever you think they fit best. But in any case, it should be possible and easy for a reader to find the results for each of the hypotheses in a section that has a heading that matches that hypothesis.

9. It would be helpful for readers to know that the feelings items were aggregated in Small et al., that feelings were assessed in Study 1 but not Study 3 (if I read their paper correctly), and that your aggregate combines the same items that were combined in Small et al. Study 1 (if that is correct). I would also like you to report the Cronbach's alpha for the aggregate (the original authors did), and if it is low, to note that (but it was high in the original, so hopefully it will be high for you, too). In addition, Table 7 needs more information - the table note should specify which variables are included in this aggregate, whether this matches the set of variables used to create the feelings aggregate in Small et al. Study 1, and report the Cronbach's alpha in your study.
10. Table 6 was hard to interpret - add a table note explaining that these are means (correct?), and that the values in from the original (in parentheses) are actual dollars donated, whereas the values in the current study are only hypothetical.
11. For the disaggregated feelings results (pp. 29-30), it is important to remind readers that the perceived impact item was not measured in Small et al., and was not included in the aggregate. Once again, I question whether the full 2x3 analysis is the best analysis - personally, I am much more curious about more specific comparisons of cells that map onto a more conceptual research questions/hypothesis.
12. The results for the perceived impact analysis need to be described in greater detail, and perhaps other analyses should be reported in addition to the main effect of identifiability when all six cells are included. At a minimum, I would like you to add a table with the descriptive results for the perceived impact variable separately for all six cells (like Table 7, but for the perceived impact variable) so that readers can see what cells might be driving the effects found in the 2x3 analyses. Then, I would like you to consider adding more discussion of what this effect looks when comparing specific subsets of cells. Without knowing what these results look like, I'm not sure what, if anything, there is to say about this, but to me the "cleanest" test of the effect of identifiability on perceived impact would be the comparison between the identifiable and statistical cells in the control level of the intervention variable (though I could also see an argument for examining all three cells in the control level). But there might be other interesting patterns, though of course you want to be careful about over-interpreting the results. Still, I think some discussion of/attention to what cells seem to be driving the effect is warranted, given that this result is messy and somewhat hard to interpret

when both intervention conditions are included together. One reason I am urging more exploration of the pattern of results here is because this result is quite critical for your overall conclusions. If it weren't for this result, it would be more or less accurate to say that your study basically found no evidence for any effect of either of the manipulated variables, and indeed quite strong evidence against a causal effect of these variables in some cases, right? However, the results for perceived impact are one indication that these null results may fail to pick up on a relevant phenomenon - maybe the identifiable victim effect is not observable in hypothetical donations, but if it affects perceived impact, and the effect is substantial, then it is reasonable to presume that it could have downstream effects. Thus, this result makes your interpretation/discussion/conclusions much more complicated (and probably more interesting to readers who want to believe the IVE, and/or to those of us who have strong priors that the identifiability manipulation must affect *something*). I admit that the lack of effect on most of the feelings variables

13. When describing the results for the associations between aggregated feelings and hypothetical donations (p. 30), please remind readers that your aggregated feelings variable included the same items as the original (if true) and therefore included all of the measured feeling variables except perceived impact (again, if true). Additionally, readers may be curious whether the perceived impact ratings correlated with hypothetical donations. I'm reluctant to ask you to add more analyses to the main text, but maybe a one or two sentence description of this result would be good to add.
14. In Table 12, I was surprised to see that you report the results for hypotheses 5a and 5b separately for each feeling variable. The original only reported the results for the aggregated feeling variable, right? It seems better to stick to the parallel analysis here (and also, this would allow you to include the result of the original study in this table). Also, because the direction of the effect for moral responsibility and donation appropriateness were never described in the text, a reader cannot evaluate whether these results support the hypothesis. Moreover, the results of the Bayesian statistical tests did not support the hypotheses for these variables, so it's a bit misleading to conclude "supported" (though it helps that the name of that column is NHST summary, but I forgot that between the time I read the table and the time I went to type this up (less than a minute later). Related to that, I wondered whether Table 12 should include a column that reports the Bayesian results. I would recommend reporting this qualitatively, even though there is a loss of information as you suggest (e.g., "strong in favor of null"), because the effect size is not intuitively interpretable, at least to me, and the NHST results is not very informative with your sample size. Ideally, you would also flag when the result is statistically significant/shows evidence in favor of the alternative,

but is in the wrong direction or has the wrong pattern, if that is the case for any of your results.

Smaller points:
It is important to report deviations from the preregistration in the main text, not just the supplement. You say that you report them in the supplement, but I wonder if you actually did report at least some of them in the main text (for example see my next point about exclusions).
It was not clear to me why you focused on the full sample for the analyses reported in the main manuscript, if you preregistered exclusion criteria.

When introducing Bayes Factors on page 7, you describe the conceptual difference between BF01 and BF10, and you lead readers to expect that you might use both and switch between them, which worried me. However, from my reading, I think you only use BF01, is that right? If not, I think you should stick to only one. Then, I think you should make that clear where you introduce BFs on page 7 (i.e., explicitly say that you'll only report BF01, and therefore that numbers greater than 1 are evidence directionally in favor of the null). You could also add, in your table notes where you define BF01, an explicit statement that values above 1 are directionally in favor of the null hypothesis.

I'm not sure you need Table 9 - two of the three results are reported in the text (and you could easily add the third), and without having the means handy it's hard to remember how to interpret the direction of these differences, so the description in text is more useful than the table anyway. (But note that as per my comments above, I am recommending that these analyses be replaced by analyses that include only the three cells in the control condition of the intervention variable.)

p. 2 "Our first goal was to conduct independent […] replication" should be "an independent […] replication"?
p. 3 - I did not understand "examine the relationship between affective feelings on the studied phenomenon" when I read it. Maybe change it to "examine effects of the the manipulations on affective feelings". It would also help to explicitly state that the first goal is looking at donations as the outcome, and then you could be more explicit here in describing the second goal as examining feelings as the outcome.
p. 3 "pay more attention to proportions or percentages" add "than to absolute numbers"
p. 3 "a higher proportion of lives seems" should be "lives saved seems"
p. 5 "three largest powered" should be "three highest-powered"
p. 5 "data show effect that" should be "effects"
p. 7 "The right panel shows the regression line" I think this should be "left panel"
p. 10 "than showed the effect" should be "that showed the effect"

p. 10 “for behavior and seem higher up” should be “for behavior and intent seems higher up” but see my point above about cutting language related to intent
p. 11: “the likelihood of original” should be “the original”
p. 12: “an implicit manner in suppressing” change “in” to “of”
p. 12: “results in the original studies” change “in” to “of”
p. 13: I didn’t understand why you included “to only one group” in the parenthetical beginning “(rather than explaining…” - the implicit intervention was still to only one group (one level of a three-level factor). I think this section will be rewritten anyway, though, when you provide a summary of the designs of the Small et al. studies as requested above.
p. 13: in the sentence beginning “In other words, our studies is a 2 x 3 design…” I would switch the order of “statistical” and “identifiable” to make the order consistent with how you present the levels in the tables (or, in any case, pick an order and be consistent throughout).
p. 13: “testthe” should be “test the”
p. 15: Table 1 note “would not be presented with” should be "would not necessarily be presented with"p. 16: “See section’ deviations…” should be “See section 'deviations…”
p. 16: “alpha level of .05” two-tailed or one-tailed?
p. 18: “between subjects’ design” delete the apostrophe
p. 19: “were attentive” should be “were inattentive”
p. 19: “and the perceived impact of donation (extension)” it’s not clear that “(extension)” refers only to the last variable, because of the “and” earlier in the sentence.
p. 24: “and their interaction” should be “or their interaction”
p. 26: “joined” should be “joint” (also check for this elsewhere)
p. 29: “segregated” should be “disaggregated” (check for this elsewhere)

In summary, I think this is a very strong manuscript and, I hope you will revise it and resubmit it to Collabra: Psychology. I look forward to receiving your revision, and if you can address the points I raised and I don’t find any new significant errors/problems, I expect to accept the next submission. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Simine Vazire
Editor in Chief
Collabra: Psychology

Author Response
Oct 6, 2023

Reply to Collabra decision letter: Small et al. (2007) replication and extension

We are grateful for the constructive suggestions for improvement. We incorporated the majority of these suggestions, resulting in what we believe is an improved version of the manuscript.
A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/hrPQqkSTBJpt

Below is a point-by-point response to the comments. The editor's comments are in bold with our reply underneath in normal script.

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
General	We removed any references to 'intent to donate' and instead reframed our dependent variable as hypothetical donations throughout the manuscript.
Introduction	We clarified several points regarding the Study of Small et al. based on your comments.We also changed Tables 2 and 8 to explain the different cells of our design and then describe tests conducted later in terms of these cells. We believe this makes it much clearer, which cells are compared in which statistical tests. We further made a lot of changes throughout to improve the writing for this final submission.
Results	We added several additional analyses to exactly match the analyses conducted by Small et al. (2007). We also restructured the results section to make it more easy to identify which hypotheses are being tested. We further added information about the relationship between perceived impact of donation and hypothetical donations. We also made several smaller changes based on the editors comments.

Reply to Editor: Prof. Simine Vazire

I apologize for the delay in getting this decision to you. Although your manuscript was streamlined, and did not go out for external review, I needed to find time to read everything carefully. From past experience with streamlined manuscripts (many from your lab!), I know that this often takes me an entire workday, and sometimes it takes a while before I'm able to clear a large chunk of time from my calendar. I know this probably does not alleviate any frustration you might have about the delay, but I will say that digging into a streamlined manuscript is one of my favorite ways to spend my time, so I look forward to the opportunity to spend an entire day immersed in that activity. This was no exception -- your manuscript did not disappoint. It is a very rich submission, with many strong aspects, and overall I feel that it is very close to publishable. Nevertheless, there are some revisions that I am asking for before accepting your manuscript. I hope you will agree that they will improve the manuscript, and I believe they are all fairly straightforward to implement. Thus, I am happy to commit to accepting your manuscript if you are able to address these points, and if I find no or few new problems/errors in your revised submission.

Thank you very much for taking the time to review our manuscript carefully and the detailed, positive, and constructive feedback.

I've numbered my points below, but many of them are interrelated and there is some repitition. I apologize for this - I wrote this decision letter as I was reading the manuscript, and tried to go back and make it more coherent, but there is still some redundancy, and the order of the points is relatively meaningless (they are not ordered from most to least important).

1. In some places you refer to your outcome measure for donations as measuring "intentions" or "intended donations", but I don't think that's accurate. From my reading of the DV, these are not intentions to donate, just hypothetical donations. That undermines the strength of some of your arguments about the far replication still speaking to the effect in the original. To be clear, I agree that these results, with this hypothetical donation DV, are still relevant and worth publishing, and that the community should care about them and their implications for the original effect. And since you describe the study as a far replication, there isn't much change needed in the framing.

However, I think you do need to change your language where you refer to intentions or a similar concept, and you need to tone down some of the arguments you make that depend on this conceptualization of your DV. After having read the reviews and response to the reviews from JDM, I want to add that I think there is probably an interesting discussion to be had about what the relationship is between answers to hypothetical scenarios, and intentions. I'd be open to you making an argument that there is a connection, but at least to me, as someone outside of JDM who interprets "intent" in the common sense way, my answer to a hypothetical scenario is not an intention, and I think that perspective should be represented in your framing.

Thank you for the feedback.
That is an interesting observation, and we see the value in being more accurate and having a better alignment between what was measured and the claims. The literature on the topic does seem rather vague about these links, and sometimes takes hypothetical donations as some proxy for intent. This does seem like something that the literature would need to resolve, yet we do not think it is important for us to tackle this given our framing as a “far replication”. There is already plenty that we are tackling here, so there is no need to also try and tackle that.

We changed the language to refrain from intent and instead refer to hypothetical donations. We went over the manuscript and tried to further tone down any claims that may suggest anything beyond the context of hypothetical donations.

We see this as the main paragraph addressing the issue (in the beginning of the discussion section):

> We caution that our results should not be considered a 'final word' on this effect but rather a motivation for future replication efforts in the form of high-powered registered reports examining hypothetical donations, donation intent, real money donations, and associated perceptions such as perceived impact. In addition, we see many promising theoretical directions for further work in this area and possibilities for rethinking and reframing the original theory.

In addition, we include two paragraphs discussing the topic in the limitations section (p.40)

2. It wasn't clear to me from the study description what participants in the implicit learning/joint statistical and identifiable information condition were asked to do. Were they only asked to donate to the identifiable victim? Were they given a choice about which victim(s) they could donate to? I found the answer to this question (only asked to donate to the identifiable victim) on page 19, but this information is necessary to interpret the research questions and hypotheses, so it needs to be explained in greater detail when the study design is described (pp. 12-15), and perhaps even when describing the results of Small et al. (pp. 11-12), because it's hard to make sense of this condition without knowing what the donation target was.

Thank you for pointing this out. We agree this needs to be clarified earlier and better.
We therefore added the following to the introduction under subsection "Small et al. (2007): Hypotheses and Findings":

> "In Study 3, Small et al. (2007) further studied the effect of implicit learning by adding a third identifiability condition, a joint condition (also referred to as "implicit learning condition") that included both a picture of the single victim and general victim statistics, resulting in a three conditions design (identifiable vs. statistical vs. joint). The donation in this joint condition was intended for the described identified victim. The presentation of victim statistics was meant to implicitly eliminate the identifiable victim effect in the joint condition arguably because providing statistics alongside the victim reminds the potential donor of the many people who would not receive help. Study 3 did not investigate how feelings predicted donations. In summary, the Study 3 design included one factor with three levels/conditions: identifiable victim, statistical victims, and the joint/implicit learning condition."

3. Related to this, I think it would help a lot if, when describing the Small et al. hypotheses and findings (pp. 11-12), you explicitly describe the design of each of the two studies. I had to go look these up to understand how your 2 x 3 design maps onto their designs, and I believe their designs are: Study 1 is a 2 (identifiability: identifiable vs. statistical) x 2 (explicit learning: intervention vs. control) between-subjects factorial design, and Study 3 has only one factor with three levels/conditions: identifiable victim, statistical victims, and the joint/implicit learning condition where the donation target is an identifiable victim but statistical information is also provided.

Thank you, that is very valuable feedback. We rewrote the description of Small et al. (2007) to explicitly describe the design of the two studies:

> Small et al. (2007) proposed that thinking analytically about the value of lives reduced giving to an identifiable victim but not to statistical victims. They also suggested that implicitly inducing analytical reasoning about the value of lives reduced donations to an identifiable victim but not to statistical victims. They conducted four experiments, and the current replication focused on Studies 1 and 3.
>
> **Study 1 design and findings**
>
> In Study 1, participants were randomly assigned to one of two conditions, with the intervention group learning about the identifiable victim effect from previous research (explicit learning condition), whereas another served as a control group. They were further randomly assigned to either the statistical victim condition, in which they read information either about the problem of starvation in different African countries, or to the identifiable victim condition, in which they received a brief description of an African girl from the Save the Children website. They were then instructed to donate any five one-dollar bills received earlier from a survey to victims they had read about in the letter. After their donation, participants rated different affective reactions they experienced towards the described victim(s). These items included feeling upset, touched, sympathetic, and morally responsible, as well as the perceived appropriateness of donating to help the described victims.
>
> To summarize, their Study 1 design was a 2 (Identifiability: identifiable vs. statistical) x 2 (Explicit Learning: intervention vs. control) between-subjects factorial design. Their results showed that in the control condition without the intervention, donations to the identifiable victim were higher than donations to statistical victims. However, the pattern was different for the participants who were assigned to the explicit learning intervention conditions and learned of the identifiable victim effect before asking to donate, with the donations being similar towards the identifiable victim compared to towards statistical victims. The explicit learning intervention, therefore, seemed to have eliminated the additional donations given towards an identifiable victim . In addition, they showed that aggregated feelings predicted donation behavior better in the identifiable victim/no intervention condition than in the other conditions.
>
> **Study 3 design and findings**
>
> In Study 3, Small et al. (2007) further studied the effect of implicit learning by adding a third identifiability condition, a joint condition (also referred to as "implicit learning condition") that included both a picture of the single victim and general victim statistics,

resulting in a three conditions design (identifiable vs. statistical vs. joint). The donation in this joint condition was intended for the described identified victim. The presentation of victim statistics was meant to implicitly eliminate the identifiable victim effect in the joint condition arguably because providing statistics alongside the victim reminds the potential donor of the many people who would not receive help. Study 3 did not investigate how feelings predicted donations. In summary, the Study 3 design included one factor with three levels/conditions: identifiable victim, statistical victims, and the joint/implicit learning condition.
Small et al. (2007) found support for implicit learning, as donations to the identified victim were lower in the joint condition compared to the identifiable victim condition.

4. One of my biggest struggles throughout the paper was figuring out which cells in your design were compared to test each hypothesis. In Table 1, you describe the hypotheses verbally, but it wasn't clear to me how this mapped onto the different cells in your design. What would help is if you would move Table 3 earlier, remove the text about hypothetical donations and affective feelings that is currently in the middle of the table (this text is misleading anyway, as it seems like the DVs are paired with the rows of IV2, when in fact all DVs were measured in all conditions), and instead fill in the table with a label for each of the six cells (e.g., "A1, A2, A3" in the top row and "B1, B2, B3" in the bottom row, or even just the letters A through F, or whatever you think is best). Then. in the table describing your hypotheses (what is currently Table 1), refer to these cells explicitly (in addition to the verbal description of the hypothesis). For main effects, this might be something like "comparing cells A1 and A2 vs. cells B1 and B2", and for an interaction this might be something like "the difference between cell A1 and B1 will be greater than the difference between cells A2 and B2".

That is very valuable feedback, and is very important for clarity and allowing reviewers and readers to follow what we have done given the integrated design. We agree and have worked to improve on this throughout the manuscript.
We updated this in the main Table 2 “Replication and extension: Summary of hypotheses” which now has a column labeled “Conditions comparisons for hypotheses”. We did the same with Tables 8 and 9 that now reference both the hypotheses and the conditions for each of the analyses.
We also reorganized the results section to better mirror the tables and explicitly refer to the tested hypotheses.

5. Related to this, you refer to the two factors as "identifiability" and "intervention" but I think this isn't quite correct. One factor is just whether there was the explicit intervention (explicit learning vs. control). But the other factor is a combination of the simple identifiable vs. statistical identifiability variable, and the implicit learning intervention, which happens to be carried out by combining the identifiable and statistical information, but is conceptually an intervention, not a level of identifiability. This makes your design quite messy,

because the two factors of the design don't map on neatly to two conceptual (causal) variables. I'm not sure how best to label the two variables given this situation. I am open to you keeping the oversimplified "intervention" and "identifiability labels", but it would help if you explicitly acknowledged that one of the identifiability levels is the implicit learning intervention in Small et al.'s Study 3 (I think). In addition, I suggest changing the "intervention" label to "explicit learning intervention" or something like that.

Thank you. This is very helpful, and we agree, this was confusing. We originally use the labels by the target article, but can see it would be helpful to improve on that.
We kept the identifiability label, and renamed the "intervention" label to "Explicit learning" throughout the manuscript. We also refer to the joint condition as "Implicit learning".
In our main hypotheses table (Table 2) and results (Tables 8 and 9) we now also make it clearer what each of those labels refers to in terms of both the specific hypothesis, and in terms of the cells compared in the unified design.

6. In the results section, you say that you followed (and extended) the analyses conducted by the target article, but because your design was different (2 x 3, which meant that you ended up crossing the explicit learning conditions with the implicit learning conditions, which I believe Small et al. never did, since those two manipulations were in separate experiments?), it is not clear what the most similar analyses would be. This becomes a bit clearer when you report the analyses for each hypothesis (starting on page 23), but even then it is not totally clear. For example, on page 23 you describe a 2 x 2, and I can guess which of the six cells got dropped (the two cells in the joint level of identifiability), but you never say this explicitly (and indeed, your Figure 2 includes all of the cells, unlike the analysis, which is only explained indirectly - it needs to be more explicit). If you alert readers early on (in the intro around pp. 12-15, where I am asking you to add a table with your design and labels for each cell) that your design makes it tricky to map each cell onto the similar/identical cell in the original studies, and then remind readers at the beginning of the results section of this challenge, and of the fact that you will need to compare subsets of cells (not the whole 2x3 design) for many of your analyses, I think this will help a lot. Then, when you get to each specific analysis (p. 23 onwards), you can once again use the labels I've asked you to create for each cell to make it super clear to readers which cells are being compared or dropped in each analysis.

We appreciate this feedback. We agree, the unified design has made it a bit more challenging to articulate how our design relates to the target's studies.
We made many changes to the manuscript to address these suggestions. In addition to re-labeling, and being clear about hypotheses and which cells are contrasted, we also added clearer indications of how each hypothesis maps onto the target article, labeling each hypothesis and

analysis as either a replication of Study 1, a replication of Study 3, or an extension that the unified design allowed us to test (see labels [S1] [S3] [E] in the different lines).

Related to this, I could not make sense of Table 8. I admit that if I tried harder, I probably could get there, but I don't think many readers are going to try harder than I have. I think adding the labels for each cell, and then explicitly saying which cells are being compared to which for each analysis, will help a lot with Table 8.

Yes, we understand and agree. We revamped Table 8 as a supplementary to the revamped Table 2, and also did similar changes to Table 9.
In all tables, we now refer to the specific hypotheses tested, and which cells are contrasted. We also clarified these better in the accompanying text.

7. Separate from the issue of being clearer about which cells are included in which analysis, I am concerned that your analyses, if I understand them correctly, did not always use the "right" cells, if the goal is to match the analyses in Small et al. For example, for the bottom half of the first row titled "Identifiability (with joint condition) [Study 3]", it doesn't sound to me like these analyses match the Small et al. Study 3 analyses, because Study 3 did not include the explicit learning intervention, right? From the text on page 24, it sounds like this analysis includes all six cells, rather than only including the three cells in the control level of the intervention variable, which is what you'd need to do to match the Small et al. Study 3 test of a main effect of identifiability. From looking at the descriptive results, I believe that when you limit the test of the effect of identifiability to include only the cells in the control condition of the intervention variable, you will find even smaller effects/stronger evidence for the null. This is important - to the extent that there is any effect of the joint condition, it seems to be driven by the cells in the intervention condition, which means that presenting the joint information only had an effect (to the extent that it had an effect at all, which is very small) when presented together with the explicit learning information. I realize the interaction was not significant, but my main point is that I think the appropriate test of the effect of implicit learning/joint presentation is a simple 3-cell comparison of the control condition cells, and this will likely show strong evidence for the null. I suspect that a similar decision (about how to alter the analyses to best match the original analyses and the conceptual question) will come up for other hypotheses, but I haven't considered/thought through each of them carefully, in part because the current structure of the results section makes it hard for me to think through what each hypothesis is and what the best test would be (see my next point). These issues also apply to the analyses of the feelings aggregate, but here the situation is more complicated because, if I understand correctly, feelings were

only measured in Study 1 in Small et al., so the most appropriate test if we want to compare the results to the original would be excluding the two cells in the joint level of the identifiability variable. However, it might be interesting to readers to know how the joint condition affected feelings, so it might also be worth doing the three-cell comparison of feelings across the three cells in the control level of the intervention variable, which would also have the advantage of matching the analyses you'd be conducting for the donation DV. On the other hand, the descriptive results in Table 7 are enough for curious readers to look at this if they want to (especially since your data will be publicly available), so perhaps you could just stick to the 2x2 that matches the analysis of feelings in Small et al., Study 1. If changing some of the analyses to better match the original studies leads to a deviation from your preregistered analyses, you will need to decide whether/where to present the preregistered analyses that I am asking you to replace. My strong preference would be that they be moved out of the main text, so that the main text only includes the most appropriate test of each hypothesis (i.e., the test that most closely matches the original), even if these were not preregistered. But it should also be very clear to readers, without having to read the supplement, what was and what was not preregistered. My reason for requesting that you change the analyses to better match the original analyses is not just out of a desire to match the original studies. I also think these tests are the right ones to test the conceptual hypotheses. That is, for example, if I want to know if the effect of the joint condition, I think the implied research question here is, "in the absence of any other interventions, does presenting the statistical information in addition to the identifiable information impact donations?". Thus, only the cells in the "control" level of the (explicit) intervention condition are relevant. This is what I assume most readers would be interested in, and what Small et al.'s Study 3 tested. Moreover, it's confusing to me that for the main effect of identifiability, you did drop some cells to make the analysis match Small et al.'s Study 1 (i.e., you dropped the two joint condition cells), but for the effect of identifiability, you didn't drop the cells that should be dropped to match Small et al.'s Study 3 (i.e., the explicit intervention condition cells). The fact that you are inconsistent about this is a further reason to revise your approach, and make it more consistent throughout, and consistent with your stated aim of matching the analyses of the original paper.

Thank you, great feedback. We realized that we should have been much clearer about what analyses match which of the target article's studies, and what analyses are extensions.
We made the necessary changes in Tables 2, 8, and 9, and in the accompanying text. We now also explicitly label each analysis, and we mirror all analyses in the target article, as well as the additional extension analyses that the unified design allowed us to execute. We are also clearer on what analyses were not pre-registered and we therefore labeled "exploratory", even if they were in the target article.

Specifically, regarding the case of the implicit learning hypothesis, we added the following hypothesis to Table 2, which matched the analyses by Small et al:

H4 (Identifiability with Implicit Learning main effect, without Explicit) [S3] *

We wrote the following in the results section:

> **H4: Identifiability with Implicit Learning (joint condition) main effect (without Explicit Learning) [Replication]**
>
> We conducted the analyses mirroring the analyses of Study 3 in Small et al (2007), without including the explicit learning intervention conditions (i.e., H4: Identifiable-Control > Joint-Control ~= Joint-Control). Although this was conducted by the target, it was not included in the pre-registration, which was focused on the unified design and included the explicit conditions (see below). We therefore labeled this analysis exploratory. We found no support for an implicit learning effect, $F(2, 499) = 0.61$, $p = .541$, $\eta_p^2 = .002$, 95% CI [.00, .02], and with strong evidence against the effect in a complementary Bayesian analysis (BF01 = 25.03). Therefore, we did not conduct any follow-up tests comparing differences between specific cells.

8. Another issue is that the results section lacks a clear structure. For example, the heading on page 23 is about an interaction effect, but the first paragraph is about a main effect (mostly, it also talks about some things that are not specific to the main effect, such as the figure and the Bayesian analytic approach). I would like you to structure the results section to match the structure of the hypotheses presented in Table 1, with a heading or subheading for each hypothesis, and text under each heading that matches the heading. I realize that some things you want to report in the results section are not specific to one hypothesis (e.g., some figures, and the description of the analytic approach, etc.) - these can be presented at the beginning, before the results speaking to each specific hypothesis, or wherever you think they fit best. But in any case, it should be possible and easy for a reader to find the results for each of the hypotheses in a section that has a heading that matches that hypothesis.

We understand and agree.
We now structure the results section according to the DVs and the hypotheses:

- Results
    - Descriptive Statistics
    - Hypothetical Donations
        - H1a, H2a, and H3a: Identifiability and Explicit Learning Main Effects and Interaction (without joint condition) [Replication]
        - H1b, H2b and H3b: Identifiability and Explicit Learning Main Effects and Interaction (with Joint Condition) [Extension]
        - H4: Identifiability with Implicit Learning (joint condition) main effect (with and without Explicit Learning) [Replication]
    - Feelings
        - H5a/b/c: Identifiability and Explicit Learning Interaction without joint condition on aggregated feelings [Replication]
        - H5d: Identifiability and Explicit Learning Interaction with joint condition on aggregated feelings [Extension]
        - Exploratory: Identifiability and Explicit Learning interaction on singular feelings
        - Associations between Aggregated Feelings and Hypothetical Donations

9. It would be helpful for readers to know that the feelings items were aggregated in Small et al., that feelings were assessed in Study 1 but not Study 3 (if I read their paper correctly), and that your aggregate combines the same items that were combined in Small et al. Study 1 (if that is correct). I would also like you to report the Cronbach's alpha for the aggregate (the original authors did), and if it is low, to note that (but it was high in the original, so hopefully it will be high for you, too). In addition, Table 7 needs more information - the table note should specify which variables are included in this aggregate, whether this matches the set of variables used to create the feelings aggregate in Small et al. Study 1, and report the Cronbach's alpha in your study.

Thank you, we followed and implemented all your suggestions.
We now report the reliability at the beginning of the "Feelings" subsection in the Results, which was 0.90. We also make it clearer which items were included in the aggregated feelings measure:

> The Cronbach's alpha for the feelings variables was 0.90. We therefore followed the methodology by Small et al (2007) and aggregated the five feelings into a single measure of aggregated feelings, combining: 1) feeling upset, 2) feeling sympathetic towards the victim(s), 3) feeling touched by the situation, 4) feeling morally responsible, and 5) feeling that it is appropriate to donate to the cause.

We now also explain which study included the aggregated feelings, how the analyses relate to what is in the target article and what was pre-registered, also updated in Tables 2 (hypotheses) and 9 (results):

> **H5a/b/c: Identifiability and Explicit Learning Main Effects and Interaction on aggregated feelings (without joint condition) [Replication]**
>
> In Small et al. (2007), aggregated feelings were measured and analyzed in Study 1, and the joint condition was introduced in Study 3. We therefore first conducted a matched analysis to their Study 1 without the joint condition. We note that our pre-registration originally focused on the analyses that included the joint condition, yet deviated from the target's Study 1, which is reported in the following section.
>
> [...]
> **H5d: Identifiability and Explicit Learning Main Effects and Interaction on aggregated feelings (with joint condition) [Extension]**
> We also ran a pre-registered extension analysis with the joint condition that went beyond the target article's Study 1. We conducted a 2 (Explicit Learning) × 3 (Identifiability) two-way ANOVA on aggregated feelings (mean across all feelings measures apart from 'perceived impact'). Results were similar to those without the joint condition, with no support for the main effects of Identifiability, the main effect of Explicit Learning, or the interaction on aggregated feelings. The Bayesian analysis further suggests evidence against an effect. Thus, we concluded that H5a/b/c and H5d were not supported.

10. Table 6 was hard to interpret - add a table note explaining that these are means (correct?), and that the values in from the original (in parentheses) are

actual dollars donated, whereas the values in the current study are only hypothetical.

We added a table note to explain this:

> *Note*.Statistics are presented in the order of Mean [Standard deviation] (condition sample size). We reported the same information for the non-aggregated feelings in the supplementary materials. Aggregated feelings were calculated following the approach by Small et al. (2007): Upset, sympathetic, touched, responsible, and appropriateness. The Cronbach's alpha for the five feelings measures was 0.90.

11. For the disaggregated feelings results (pp. 29-30), it is important to remind readers that the perceived impact item was not measured in Small et al., and was not included in the aggregate. Once again, I question whether the full 2x3 analysis is the best analysis - personally, I am much more curious about more specific comparisons of cells that map onto a more conceptual research questions/hypothesis.

We believe that follow up tests of specific cells are better conducted if we find evidence for a difference in the overall ANOVA tests (as we did for the perceived impact). As we do not find compelling evidence for any effect in the 2 x 3 ANOVAs we did not conduct any follow ups. As per conceptual research questions, we think the most interesting analysis is the correlation between aggregated feelings and hypothetical donations. Here Small et al. (2007) reported that this correlation is stronger in the identifiability/no intervention condition, which is seen as a potential explanation of the effect. We do not find support for this pattern. These results are presented in the next section.

We now explicitly state in both text and tables which feelings are included in the aggregated feelings measure. We also separated perceived impact to its own section.

12. The results for the perceived impact analysis need to be described in greater detail, and perhaps other analyses should be reported in addition to the main effect of identifiability when all six cells are included. At a minimum, I would like you to add a table with the descriptive results for the perceived impact variable separately for all six cells (like Table 7, but for the perceived impact variable) so that readers can see what cells might be driving the effects found in the 2x3 analyses. Then, I would like you to consider adding more discussion of what this effect looks when comparing specific subsets of cells.

We reported all perceived impact descriptives in Table 7. We follow up with pairwise comparisons along the Identifiability factor as only this factor showed a main effect and there was no interaction.

Table 7

Perceived Impact (extension): Descriptives

	Identifiable victim	Statistical victim	Joint	Total
Explicit learning intervention	3.47 [1.23] (170)	2.91 [1.37] (159)	2.94 [1.35] (173)	3.11 [1.34] (502)
No explicit learning intervention	3.31 [1.25] (165)	2.99 [1.34] (176)	3.11 [1.42] (161)	3.14 [1.34] (502)
Total	3.39 [1.24] (335)	2.96 [1.35] (335)	3.02 [1.39] (334)	3.12 [1.34] (1004)

Note. Statistics are presented in the order of Mean [Standard deviation] (condition sample size).

Without knowing what these results look like, I'm not sure what, if anything, there is to say about this, but to me the "cleanest" test of the effect of identifiability on perceived impact would be the comparison between the identifiable and statistical cells in the control level of the intervention variable (though I could also see an argument for examining all three cells in the control level). But there might be other interesting patterns, though of course you want to be careful about over-interpreting the results. Still, I think some discussion of/attention to what cells seem to be driving the effect is warranted, given that this result is messy and somewhat hard to interpret when both intervention conditions are included together.

We did not see much need to go into the details of the contrasts between the conditions, the Explicit Learning main effect, or the interaction, given that our ANOVA model only showed support for a main effect of identifiability. We believe that looking at Table 7 also makes that pattern fairly clear, together with our explanation:

H6: Effect of Identifiability and Explicit Learning on Perceived Impact of Donation

> We ran a 2 (Explicit Learning) x 3 (Identifiability) two-way ANOVA to determine how the perceived impact of the donation differed depending on these two factors. We find evidence for an effect of Identifiability on perceived impact, $F(2, 998) = 10.5$, $p = .00003$, $\eta_{(p)}^2 = .021$ [.006, .040], BF01 = 0.003. However, we do not find evidence for an effect of Explicit Learning on perceived impact, $F(1, 998) = 0.11$, $p = 0.74$, $\eta_{(p)}^2 = .000$ [.000, .004], BF01 = 13.69, or for an interaction between Explicit Learning and Identifiability, $F(2, 998) = 1.48$, $p = 0.229$, $\eta_{(p)}^2 = .002$ [.000, .012], BF01 = 11.05.
>
> As we only found evidence for an effect of Identifiability, we follow up with post-hoc tests of this factor. We found support for higher perceived impact in the identifiable victim condition (M = 3.39, SD = 1.24) compared to both the statistical victim (M = 2.96, SD = 1.35), $t(668) = 4.35$, $p < .001$, $d = 0.34$, 95% CI [0.18, 0.49], and joint condition (M = 3.02, SD = 1.39), $t(667) = 3.67$, $p < .001$, $d = 0.28$, 95% CI [0.13, 0.44]. Both of these results also held up in the Bayesian analysis (BF01 = 0.001 and 0.02).

One reason I am urging more exploration of the pattern of results here is because this result is quite critical for your overall conclusions. If it weren't for this result, it would be more or less accurate to say that your study basically found no evidence for any effect of either of the manipulated variables, and indeed quite strong evidence against a causal effect of these variables in some cases, right? However, the results for perceived impact are one indication that these null results may fail to pick up on a relevant phenomenon - maybe the identifiable victim effect is not observable in hypothetical donations, but if it affects perceived impact, and the effect is substantial, then it is reasonable to presume that it could have downstream effects. Thus, this result makes your interpretation/discussion/conclusions much more complicated (and probably more interesting to readers who want to believe the IVE, and/or to those of us who have strong priors that the identifiability manipulation must affect something). I admit that the lack of effect on most of the feelings variables

We agree this is complex, and we do not have enough information to understand why the perceived impact effect did not translate to an effect regarding hypothetical donations. We feel it important to be very cautious in interpreting these effects, and to make any assumptions about what might be reasonable regarding downstream effect from perceived impact to hypothetical versus real donations.

We think it best to focus on reporting the results as is, and allowing for follow-up research to try and look further to try and better explain what is going on here. In our discussion we noted the following:

> In our extension adding a measure of perceived impact, we found support for perceived impact of the hypothetical donations as higher for an identifiable victim compared to statistical victims, and with support for an association between perceived impact and hypothetical donations, though it somehow failed to translate to an effect on hypothetical donations. Further research is needed to try and understand the links between perceived impact, hypothetical donations, intent to donate, and actual donations.

(sidenote: seemed like your last sentence has been cut off)

13. When describing the results for the associations between aggregated feelings and hypothetical donations (p. 30), please remind readers that your aggregated feelings variable included the same items as the original (if true) and therefore included all of the measured feeling variables except perceived impact (again, if true).

We took the following actions:

1. The section "Associations between Aggregated Feelings and Hypothetical Donations" is now a subsection of the "Feelings" section.
2. We explicitly write that "We examined the associations between five aggregated feelings (as above) and hypothetical donations, summarized in Table 10."
3. In Table 10 we added the following note: "Aggregated feelings refer to averaging the feelings of being upset, being sympathetic, being touched, moral responsibility, and donation appropriateness into a single composite. The Cronbach's alpha for the five feelings measures was 0.90."

Additionally, readers may be curious whether the perceived impact ratings correlated with hypothetical donations. I'm reluctant to ask you to add more analyses to the main text, but maybe a one or two sentence description of this result would be good to add.

The "perceived impact" section includes the following:

> Further, the perceived impact of donation was correlated to hypothetical donations, $r(1002) = 0.54$, $p < .001$. The correlation is comparable for all cells of our design ranging from 0.48 in the identifiable victim/no intervention condition to 0.60 in the statistical victim/intervention condition (see supplementary materials for the correlation in each cell of our design).

14. In Table 12, I was surprised to see that you report the results for hypotheses 5a and 5b separately for each feeling variable. The original only reported the results for the aggregated feeling variable, right? It seems better to stick to the parallel analysis here (and also, this would allow you to include the result of the original study in this table). Also, because the direction of the effect for moral responsibility and donation appropriateness were never described in the text, a reader cannot evaluate whether these results support the hypothesis. Moreover, the results of the Bayesian statistical tests did not support the hypotheses for these variables, so it's a bit misleading to conclude "supported" (though it helps that the name of that column is NHST summary, but I forgot that between the time I read the table and the time I went to type this up (less than a minute later). Related to that, I wondered whether Table 12 should include a column that

reports the Bayesian results. I would recommend reporting this qualitatively, even though there is a loss of information as you suggest (e.g., "strong in favor of null"), because the effect size is not intuitively interpretable, at least to me, and the NHST results is not very informative with your sample size. Ideally, you would also flag when the result is statistically significant/shows evidence in favor of the alternative, but is in the wrong direction or has the wrong pattern, if that is the case for any of your results.

Thank you, very valuable feedback. We agree with all the points made.
We have made major changes to Table 11 (previously Table 12):

1. We now mirror the table to the hypotheses, focusing only on the hypotheses that relate to the replication.
2. We now report feelings on the aggregate, without the separate feelings.
3. We now report the Bayes Factor in support of the null BF_{01} in a column "BF_{01} Evidence for null".
4. We adjusted all hypotheses and the reported findings to the target's Study 1 (2 x 2, without joint conditions) and Study 1 (1 x 3, without Explicit Learning conditions) designs.
5. We also added a section with the matched analysis to the results, and are now clearer on what we pre-registered (with joint) compared to what the target article reported in Study 1 (without joint conditions).

Smaller points:

It is important to report deviations from the preregistration in the main text, not just the supplement. You say that you report them in the supplement, but I wonder if you actually did report at least some of them in the main text (for example see my next point about exclusions).

In our revision, in both the text and the tables, we worked to be more explicit about which analyses were pre-registered, and what deviations took place. We also added indications of pre-registered hypotheses and clarification notes to the tables.

It was not clear to me why you focused on the full sample for the analyses reported in the main manuscript, if you preregistered exclusion criteria.

Apologies, we wrote our exclusions section too briefly, and should have been clearer about what we did.

In our pre-registration exclusions criteria section (in the supplementary, p. 37; https://osf.io/dqch2) we wrote the following:

Generalized exclusion criteria

The default generalized exclusion criteria we use in our pre-registration is the following: "We will focus on our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we will also determine further findings reports with exclusions. In any case, we will report exclusions in detail with results for the full sample and results following exclusions (in either the manuscript or the supplementary).
General criteria:

1. Participants indicating a low proficiency of English (self-report < 4, on a 0-6 scale)
2. Participants who self-report not being serious about filling in the survey (self-report < 3, on a 0-4 scale).
3. Participants who correctly guessed the hypothesis of this study in the funnelling section.
4. Participants who have already seen or done the survey before.
5. Participants who failed to complete the survey. (duration = 0, leave question blank)
6. Participants not from the United States.

To be clearer about this, in the revision we amended the main text in the "Exclusions" subsection to the following:

> We pre-registered that "We will focus our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we will also determine further findings reports with exclusions", with several exclusions criteria for the supplementary analyses: low English proficiency (scored lower than 4 on a scale of 0 to 6); not being serious in completing the survey (scored lower than 3 on a scale of 0 to 4); correctly guessed the hypotheses; already seen the survey before; failure to complete the survey or completed in less than a minute; and not from the United States.

A related question that may arise is why we pre-registered to focus on the full sample.
The reason is that exclusions tend to be very categoric, post-hoc, and subjective (e.g., is serious = 5 really much better than serious = 4? is self-reported English proficiency = 4 really better than English proficiency = 3?). Given that, we rarely see any reason to use these arbitrary thresholds to reduce the number of participants, and typically only result to using those as supplementary analyses to try and alleviate any concerns for data quality.
Also, our experience so far has been that these exclusions tend to have very little if any impact on the results, yet have come to include those in our replication templates in an attempt to pre-empt any criticism about our sample's seriousness, comprehension, and attentiveness.

When introducing Bayes Factors on page 7, you describe the conceptual difference between BF01 and BF10, and you lead readers to expect that you might use both and switch between them, which worried me. However, from my reading, I think you only use BF01, is that right? If not, I think you should stick to only one. Then, I think you should make that clear where you introduce BFs on page 7 (i.e., explicitly say that you'll only report BF01, and therefore that numbers greater than 1 are evidence directionally in favor of the null). You

could also add, in your table notes where you define BF01, an explicit statement that values above 1 are directionally in favor of the null hypothesis.

We changed the text as requested:

> "In our paper we report BF01. In other words, Bayes factors have the null in the numerator and the alternative in the denominator, and denote evidence in favor of the null hypothesis."

And in the tables:

> "BF01 denotes the Bayes factor in favor of the null."

I'm not sure you need Table 9 - two of the three results are reported in the text (and you could easily add the third), and without having the means handy it's hard to remember how to interpret the direction of these differences, so the description in text is more useful than the table anyway. (But note that as per my comments above, I am recommending that these analyses be replaced by analyses that include only the three cells in the control condition of the intervention variable.)

We deleted Table 9 and added the third hypothesis test to the text instead.

p. 2 "Our first goal was to conduct independent [...] replication" should be "an independent [...] replication"?

Thank you. Fixed.

p. 3 - I did not understand "examine the relationship between affective feelings on the studied phenomenon" when I read it. Maybe change it to "examine effects of the the manipulations on affective feelings". It would also help to explicitly state that the first goal is looking at donations as the outcome, and then you could be more explicit here in describing the second goal as examining feelings as the outcome.

We changed it to:

> "Our second goal was to examine associations between affective feelings and hypothetical donations examining the impact of identifiability and explicit learning on affective feelings."

p. 3 "pay more attention to proportions or percentages" add "than to absolute numbers"

p. 3 "a higher proportion of lives seems" should be "lives saved seems"

p. 5 "three largest powered" should be "three highest-powered"

p. 5 "data show effect that" should be "effects"

p. 7 "The right panel shows the regression line" I think this should be "left panel"

p. 10 "than showed the effect" should be "that showed the effect"

p. 10 "for behavior and seem higher up" should be "for behavior and intent seems higher up" but see my point above about cutting language related to intent

p. 11: "the likelihood of original" should be "the original"

p. 12: "an implicit manner in suppressing" change "in" to "of"

p. 12: "results in the original studies" change "in" to "of"

We fixed these issues. Thank you for pointing them out.

p. 13: I didn't understand why you included "to only one group" in the parenthetical beginning "(rather than explaining..." - the implicit intervention was still to only one group (one level of a three-level factor). I think this section will be rewritten anyway, though, when you provide a summary of the designs of the Small et al. studies as requested above.

We removed “to only one group”

p. 13: in the sentence beginning "In other words, our studies is a 2 x 3 design..." I would switch the order of "statistical" and "identifiable" to make the order consistent with how you present the levels in the tables (or, in any case, pick an order and be consistent throughout).

We switched the order. We also changed the order of factors so that identifiability appears first, which makes more sense and is more consistent with the rest of the manuscript.

p. 13: "testthe" should be "test the"

p. 15: Table 1 note "would not be presented with" should be "would not necessarily be presented with"p. 16: "See section' deviations..." should be "See section 'deviations..."

p. 16: "alpha level of .05" two-tailed or one-tailed?

ANOVAs and F-tests are always one tail; we kept it as is, since we were worried that explicitly stating “one tail” may confuse readers as it may imply that a two sided test is possible.

p. 18: "between subjects' design" delete the apostrophe

p. 19: "were attentive" should be "were inattentive"

p. 19: "and the perceived impact of donation (extension)" it's not clear that "(extension)" refers only to the last variable, because of the "and" earlier in the sentence.

p. 24: "and their interaction" should be "or their interaction"

p. 26: "joined" should be "joint" (also check for this elsewhere)

p. 29: "segregated" should be "disaggregated" (check for this elsewhere)

We fixed these issues.

In summary, I think this is a very strong manuscript and, I hope you will revise it and resubmit it to Collabra: Psychology. I look forward to receiving your revision, and if you can address the points I raised and I don't find any new significant errors/problems, I expect to accept the next submission. Please see the instructions below for submitting your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

Thank you very much for all the detailed feedback. We are very grateful.

Editor Final Decision: Accept
Oct 15, 2023

Dear Gilad ,

I have now had a chance to read over your manuscript “Revisiting and Rethinking the Identifiable Victim Effect: Replication and Extension of Small, Loewenstein, and Slovic (2007)”, along with the letter describing the changes you made. Thank you for

your responsiveness to the concerns I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Simine Vazire
Editor in Chief
Collabra: Psychology