Perceived morality of direct versus indirect harm: Replications of the preference for indirect harm effect

Ignazio Ziano[1]

Grenoble Ecole de Management, F-38000 Grenoble, France

Yu Jie Wang[1], Sydney Susanto Sany[1]

Department of Psychology, University of Hong Kong, Hong Kong SAR

Long Ho Ngai[2], Yuk Kwan Lau[2], Iban Kaur Bhattal[2], Pui Sin Keung[2], Yan To Wong[2], Wing Zhang Tong[2]

Department of Psychology, University of Hong Kong, Hong Kong SAR

Bo Ley Cheng, Hill Yan Chan

Department of Psychology, University of Hong Kong, Hong Kong SAR

Gilad Feldman[3]

Department of Psychology, University of Hong Kong, Hong Kong SAR

Royzman and Baron (2002) demonstrated that people prefer indirect harm to direct harm: they judge actions that produce harm as a by-product to be more moral than actions that produce harm directly. In two preregistered studies, we successfully replicated Study 2 of Royzman and Baron (2002) with a Hong Kong student sample (N = 46) and an online American Mechanical Turk sample (N = 314). We found consistent evidential support for the preference for indirect harm phenomenon (d = 0.46 [0.26, 0.65] to 0.47 [0.18, 0.75]), weaker than effects reported in the original findings of the target article (d = 0.70 [0.40, 0.99]). We also successfully replicated findings regarding reasons underlying a preference for indirect harm (directness, intent, omission, probability of harm, and appearance of harm). All materials, data, and code are available at osf.io/ewq8g.

Keywords: direct harm, indirect harm, morality, pre-registered replication, preference for indirect harm

Judgments of morality do not only depend on the result of an action, but also on the way that it was performed. For instance, acts of omission are considered more moral than acts of commission, despite leading to the same result (omission bias) (Spranca, Minsk, & Baron, 1991). In their 2002 article, Royzman and Baron found that people preferred indirect harm to direct harm and considered indirect harm as more moral (Studies 1 and 2). In addition, omission bias (Jamison, Yay, & Feldman, 2020) was

found to be weaker for indirect compared to direct harm (Study 3).What is the difference between direct and indirect harm? Consider two actors, Ann and Bob, with Ann inflicting harm on Bob. An example for direct harm would be for Ann to harm Bob by pushing him off the swing. An example of indirect harm would be for Ann to saw down the tree branch to which the swing is attached, which would then in turn lead to Bob falling down and getting hurt. Both actions lead to the same outcome involving harm –

---

[1] Joint first authors

[2] Joint fourth authors

[3] Corresponding author

Bob getting hurt - yet the difference is regarding the direct link in the causal chain of events. In principle, indirect action could be performed without Bob ever being involved and, in such case, would result in no harm to Bob. Royzman and Baron (2002) hypothesized and found that even if a negative outcome is the same, people judge the morality of actions leading to that negative outcome as dependent on whether there was a direct or indirect link between the action and outcome. This in turn resulted in a strategic preference for indirect harm. In order to minimize accountability when inflicting harm, people show a preference for inflicting indirect over direct harm.

Impact of "The preference for indirect harm"

Preference for indirect harm is central in the understanding of moral judgment. In his seminal study, Milgram (1974) observed that people were more likely to commit harm if they did not have physical contact with the victim, i.e., when the harm they had to inflict to the experimenter's confederates was less direct. In general, dislike for physical contact with the victim may be caused by an overall a preference for indirect harm. Cushman, Young, and Hauser (2006) summarized and tested three principles of harm: action, intention, and contact. The second principle, which they termed the 'intention principle', is an extension to the preference for indirect harm: people prefer harm as a byproduct rather than the main goal of an action. They found corroborating evidence for indirect harm as being an intuitive guide to moral judgment, building on work by Haidt and Hersh (2001), showing that participants were unable to explain why they would prefer indirect to direct harm. Hauser, Cushman, Young, Kang-Xing Jin, and Mikhail (2007) found further support for preference for indirect harm across cultures, including that participants were unable to readily provide explanations for it. In line with these results, more recent research found further support for the intuitive nature of preference for indirect harm, as evaluation mode (joint vs. separate) moderated the effect (Paharia, Kassam, Greene, & Bazerman, 2009).

Preference for direct harm can be linked to various practices observed in everyday life. For example, Bennett (1966) compared direct and indirect action leading to the same outcome – the death of a fetus. Some Catholic hospitals – opposed to abortion on

principle – would consent to performing hysterectomy on pregnant women whose lives were in danger, while they would not consent to perform an abortion. The hysterectomy would not only kill the fetus, but also make the woman sterile. In these cases, Catholic hospitals would prefer an action that leads to a worse indirect harm (the death of the fetus and lifetime infertility for the woman) than a direct action leading to a less harm (the death of the fetus), on the religious grounds that indirect harm to the fetus is acceptable, while direct harm is not.

Choice of study for replication

We chose the Royzman and Baron (2002) study based on two factors: absence of direct replications and impact. To the best of our knowledge there are no published direct replications of this study thus far. The article has had significant impact on scholarly research in the area of moral psychology. At the time of writing, there were 173 Google Scholar citations of the article and many important follow-up theoretical and empirical articles, such as the Cushman et al. (2006) three principles of harm, and the investigation of Hauser et al. (2007) on the dissociation between the conscious nature of moral judgments (such as preference for indirect harm) and the intuitive nature of moral justifications (such as the intuition principle).

The original article consisted of three scenario-based studies using university (Study 1: N = 176) and online samples (Study 2: N = 54; Study 3: N = 69). In Studies 1 and 2 Royzman and Baron (2002) asked participants to directly compare actions that lead to the same amount of harm and the same amount of a beneficial outcome. In the first scenario of Study 2, for example, study participants had to compare the morality of two actions (action A and action B) leading to the same harmful outcome – preventing an alcoholic patient from receiving a liver transplant – either by lowering his priority on an organs transplant list (direct option) or by increasing everyone else's priorities (indirect option), by indicating whether they perceived action A or action B to be more wrong. In the present investigation, we conducted two replication attempts of the two scenarios fully detailed in Study 2 of Royzman and Baron (2002).

Original findings in target article

A summary of the findings in the target article is provided in Table 1. The preference for indirect

harm effect was d = 0.70, 95% CI [.40; 0.99], a medium to strong effect. They examined considerations and found support for all proposed mechanisms, with statistically significant correlations (r = .47 to r = .70) when participants deemed the consideration a reason for moral judgment ('predicted' column in Table 1). They found weaker and sometimes statistically non-significant correlations (r = .01 to r = .16) when participants did not deem the consideration a reason underscoring a preference for indirect harm ('opposite' column in Table 1).

Table 1

Summary of original findings in Royzman and Baron (2002)

Factors	Effect (d)	CIL	CIH
Morality	0.70	0.40	0.99
	Predicted (Direct)	Opposite (Indirect)	Morality judgment r
Probability	12.0%	4.9%	.467
Directness			
Reason	15.0%	3.5%	.649
Not a Reason	17.4%	5.1%	.099
Appearance			
Reason	15.7%	3.2%	.609
Not a Reason	27.1%	6.3%	.157
Omission			
Reason	16.0%	2.8%	.553
Not a Reason	22.2%	6.5%	.092
Intent			
Reason	16.9%	3.9%	.698
Not a Reason	32.4%	5.6%	.012

Note. Correlations with the morality question, original study, according to whether or not it was cited as a reason for a moral judgment, from Royzman and Baron (2002), p.174. 'Predicted' indicates the share of responders indicating that the direct action was more wrong; 'Opposite' indicates the share of responders indicating that the indirect action was more wrong. All correlations above .092 are significant at α = .05.

Methods

Pre-registration

In each of the replication studies, we pre-registered the experiment on the Open Science Framework and data collection was launched soon after. Pre-registrations, power analyses, disclosures, and all materials used in the experiments are available in the supplementary materials. These together with data and code were shared on the Open Science Framework (project: osf.io/ewq8g; pre-registration Hong Kong undergraduate sample: osf.io/qdn2m; pre-registration online American sample: osf.io/hwsdc).

Power analyses and deviations from power analysis preregistration

Power analyses indicated that 24 participants would be sufficient to have 95% power of detecting the original effect (d = 0.70) with a one-tailed alpha of .05, using a one-sample t-test as in the original article. The preregistration for the first data collection planned to sample 70 participants among Hong Kong University students, a decision based on convenience, as these participants were students in a Psychology course. Of that sample, we were able to collect 49 participants, given that participation was voluntary. After excluding the students who designed this very replication, 46 participants remained. Sensitivity analyses indicate that this sample size provides approximately 99.8% power to detect the original effect with a one-tailed alpha of .05.

The second online data collection on Amazon Mechanical Turk (MTurk) was part of a larger project of replications of psychology findings and this study was combined with other replications, random presentation order. The final sample size (N = 314) is due to power analyses related to the other replications running in the same data collection. Sensitivity analyses indicate 99.9%+ power to detect the original effect with a one-tailed alpha of .05.

Procedure

The first replication was considered a pre-test and conducted in an undergraduate course at a university in Hong Kong. Students worked in teams of 3 to 6 to design and run a series of replications, and one of the replications was Royzman and Baron's (2002) study. The students then served as the target sample for the experiments designed by their classmates in which they had not designed and had no

knowledge prior to participation. The course materials covered classic judgement and decision-making literature, which means that the students were made aware of a wide array of heuristics and biases, and the experiment therefore should be considered a very conservative test of the effect in a non-naive sample.

Students were randomly assigned into replication teams with different target studies for replication. Student groups designed the experiment survey, conducted effect size calculations, ran power analyses, and wrote pre-registrations. Pre-registrations on the OSF and data collections were managed by course instructor. All the students registered in the course were invited to take part as respondents in the study. To ensure anonymity, students were only asked to indicate which replication group they belonged to and those were later excluded from the data analysis of the study they designed. The final sample included the students that were not involved in planning the study, totaling 46 participants (15 males, 31 females; M_{age} = 20.2, SD_{age} = 0.99).

For the second replication, two advanced course undergraduate students unrelated to the first replication worked independently to analyze the target article. They conducted effect-size calculations, power analyses, and each separately wrote a pre-registration plan. They then reviewed each other's work and made final revisions, reviewed by the teaching assistant and the course coordinator. Both plans were pre-registered on the OSF prior to data collection by the corresponding author, who was the course instructor of the first replication and the advanced course. The final sample included 314 American MTurk workers, recruited using Turk-Prime.com (Litman, Robinson, & Abberbock, 2017) (173 males, 141 females, M_{age} = 36.8, SD_{age} = 11.3). We note that the pre-registration plans included different references to possible exclusion criteria addressing generalized factors such as seriousness, English proficiency, etc. We conducted our analyses both with and without exclusions and found that exclusions had little effect on the results. For the sake of brevity, the findings reported below are without any exclusions. A comparison of the target article sample and the replication samples is provided in Table 3. In both replication attempts, participants evaluated the two scenarios described in detail (out of the eight total scenarios; six were not described) in Royzman and Baron's (2002) Study 2, assessing participants' preference for indirect harm.

The following was the organ transplant scenario (Scenario 1 in target article):

"X is in charge of a computer database controlling the distribution of available organ transplants. The first person in line for a difficult-to-get liver transplant is Mr. Y. Mr. Y was an alcoholic and his drinking ruined his liver. Y no longer drinks. The rules say that past alcohol use should not be considered, but X still thinks that Y should not get priority, so he decides to break the rules and prevent Y from getting the next liver. He can do this in two ways:
•[Direct] X can lower Y's priority score by 20 points.
•[Indirect] X can raise everyone else's priority score by 20 points."

The following was the zoo scenario (Scenario 2 in the target article):

"A zoo has been created to conserve 200 species of wild animal that have become extinct elsewhere. The zoo is now threatened with a parasitic disease that infects the animals. X, the zookeeper, has two options:
•[Direct] Painlessly poison the animals in which the parasite reproduces, thus saving the other animals. Five species will become extinct.
•[Indirect] Poison the parasites. The same poison will cause five animal species to become extinct.

In both cases, X is sure that he will save most of the species and lose five. The five lost species are of equal value in both cases."

Measures

Morality. After each of the two scenarios, participants were asked which of the two options was morally worse (1 = *A is much more wrong*; 2 = *A is a little more wrong*; 3 = *Equal*; 4 = *B is a little more wrong*; 5 = *B is much more wrong*; note that higher scores indicate higher morality for the indirect option).

Reasons: Considerations for morality evaluations. Participants compared the two options in each scenario on five factors: Directness, intentionality, appearance, and action-omission on a five point scale (1 = factor is more applicable to the direct harm option thus the option is more immoral; 2 = direct harm option is not more immoral even though factor is more applicable to it; 3 = factor is equally applicable to both the options [equal morality]; 4 = factor is more applicable to indirect harm option, thus the

option is more immoral; 5 = indirect harm option is not more immoral even though the factor is more applicable to it), and probability of harm on a three point scale (1 = *More likely to cause harm in A than in B*; 2 = *Equally likely to cause harm in A and B*; 3 = *More likely to cause harm in B than in A*). Measures are reported in full in the supplementary.

Replications evaluation

We aimed to compare the replication effects with the original effects in the target article (d = 0.70, 95% CI [0.40; 0.99]) using two methods: (1) we categorized the comparison of effects using the criteria set by LeBel, Vanpaemel, Cheung, and Campbell (2019), and (2) we conducted equivalence testing using the TOSTER module (Lakens, Scheel, & Isager, 2018). Figures summarizing these criteria are available in the supplementary materials. Table 2 provides a classification of the replications using the criteria

LeBel, McCarthy, Earp, Elson, & Vanpaemel, (2018) criteria. We summarize the two replications as "very close replications".

Table 2
Classification of replications based on LeBel et al. (2018)

Design facet	Hong Kong replication	MTurk replication
IV operationalization	Same	Same
DV operationalization	Same	Same
IV stimuli	Same	Same
DV stimuli	Same	Same
Procedural details	Different	Different
Physical settings	Same	Same
Contextual variables	Different	Different
Replication classification	Very close replication	Very close replication

Note. Information on this classification is provided in Lebel et al. 2018. See also figure provided in the supplementary

Table 3
Difference and similarities between original studies and replication attempts

	Royzman & Baron 2002	Hong Kong undergraduate students	American MTurk workers
Sample size	54	46	314
Geographic origin	US American	Hong Kong SAR	US American
Gender	17 males, 37 females	15 males, 31 females	173 males, 141 females
Median age (years)	34	20	34
Average age (years)	Not reported	20.2	36.8
Age range (years)	17-69	19-22	21-71
Medium (location)	Computer (online)	Computer (online)	Computer (online)
Compensation	$3 (only after participant written request)	None (volunteers)	Nominal payment
Year	Not reported (during or before 2002)	2018	2018
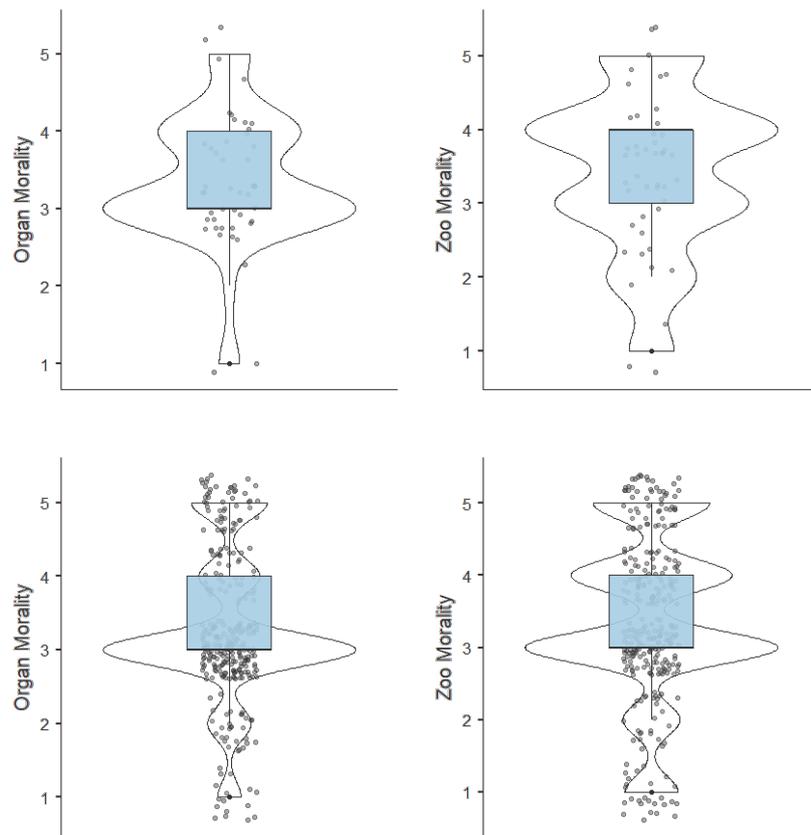
Figure 1. Plots for the morality ratings prior to categorizing. The two plots on the first row are for Hong Kong sample and the two plots on the second row are for the American sample. The first of every plot pair is for the organ scenario, and the second is for the zoo scenario. The scale is from 1 to 5, with 3 representing the mid-point. Higher values indicate higher morality ratings for the indirect option.

Results

Preference for indirect harm

Violin plots for the raw morality ratings (on a scale from 1 to 5) are provided in Figure 1. Across the two scenarios in two experiments the ratings were higher than the midpoint neutrality rating of 3.

For the analyses, we followed the method set by Royzman and Baron (2002) and recoded the morality ratings as 0 for the indifference point (3 = Equal), - 1 for the direct action being more wrong (1 = A is much more wrong; 2 = A is a little more wrong), and -1 for the indirect action being more wrong (4 = B is a little more wrong; 5 = B is much more wrong). We then ran a series of one-sample, one-sided t-tests comparing to the converted midpoint of 0, followed by dependent t-tests comparing the organ and zoo scenarios in each sample (two-sided), and finally equivalence testing comparing to the effects of the target article original findings. Note that this strategy, albeit not ideal given the low number of response categories (three) and the grouping of responses, was the one used by the original authors. We therefore complemented these analyses with non-parametric testing (Wilcoxon's signed-rank test) for the one-sample tests against the scale midpoint and for the comparison between scenarios (alpha = .05). The effect size comparisons should, however, be interpreted with caution, since the original effect size was obtained from data across eight scenario (we did not have access to the remaining six scenarios, nor to the original data, and to the effect sizes per scenario). The findings are summarized in Table 4. The findings were consistent across the two replication attempts, with similar point estimates and overlap in 95% confidence intervals. The effects in both replications were in the same direction and supported the original study's findings, but with weaker effects.

Table 4

Preference for indirect harm findings summary: Morality ratings one-sample t-tests

	M	SD	Statistic	p	d	CIL	CIH	Interpretation
Original (N=54)								
Combined effect			t(53) = 5.12	< .001	0.70	0.40	0.99	Baseline
Hong Kong (N=46)								
HK organ	.33	.60	t(45) = 3.70	< .001	0.55	0.23	0.86	Signal; consistent
			W = 198	< .001				
HK zoo	.33	.79	t(45) = 2.80	= .004	0.41	0.11	0.71	Signal; consistent
			W = 408	= .005				
Comparison			t(45) = 0.00	= 1.00	0.00	0.00	0.00	No evidence for difference
			W = 139.5	= .97				
Equivalence HK organ			t(45) = -1.03	= .154				Similar effect
Equivalence HK zoo			t(45) = -1.93	= .03				Weaker effect
MTurk (N = 314)								
MTurk organ	.15	.65	t(313) = 4.16	< .001	0.24	0.12	0.34	Signal; inconsistent, positive (weaker)
			W = 6627	< .001				
MTurk zoo	.26	.73	t(313) = 6.31	< .001	0.36	0.24	0.47	Signal; consistent
			W = 12988	< .001				
Comparison			t(313) = -2.03	= .040	0.11	.003	.21	Weak to no differences
			W 6624.5	= .066				
Equivalence MTurk organ			t(313) = -8.19	< .001				Weaker effect
Equivalence MTurk zoo			t(313) = -6.05	< .001				Weaker effect

Note. Categorized morality scores are -1 to 1, with 0 as the mid-point. Higher values indicate higher morality ratings for the indirect option. The tests are one-sample t-test comparing to 0. Comparisons are one-sided paired t-tests (alpha = .05) comparing the organ and zoo scenarios within that sample. TOST are TOSTER equivalence test analyses comparing to the effect-size found in the original findings of the target article. The interpretation column is according to the criteria set by LeBel et al. (2019) or equivalence testing (Lakens et al., 2018). "W" indicates the W statistics Wilcoxon's signed-rank non-parametric test.

Reasons: Considerations for morality evaluations

We followed the procedure in the target article to test reasons for morality evaluations and the preference for indirect harm effect by examining correlations between ratings of morality and considerations - directness, appearance, omission, and intent. Ratings were coded as either being more applicable to the direct, indirect, or neither option, and then as either being a reason or not for the morality ratings. The findings are summarized in Table 5.

We found support for the original study findings with medium to strong correlations (Hong Kong organ: r = .29 to .71; Hong Kong zoo: r = .32 to .90; MTurk organ: r = .36 to .56; MTurk zoo: r = .49 to .63) between each factor and morality ratings when the factor was indicated as a reason, and much weaker correlations, of which half were negative, contrary to predictions (Hong Kong organ: r = -.11 to .26; Hong Kong zoo: r = -.20 to .22; MTurk organ: r = .10 to .29; MTurk zoo: r = .13 to .29) when the factor was not indicated as a reason. Probability ratings were all positive and ranged from r = .14 to .50 across the samples and scenarios.

Royzman and Baron (2002) furthered add an indication to better contextualize the psychological mechanisms underlying preference for indirect harm. They classified answers to the considerations into two categories, 'predicted' and 'opposite'. 'Predicted' represented the share of responders indicating that the direct action was more wrong (thus indicating preference for *indirect* harm, in line with predictions); 'Opposite' represented the share of responders indicating that the indirect action was more wrong (thus indicating preference for *direct* harm, contrary to predictions). Royzman and Baron

(2002) further classified these answers based on whether participants find that the specific consideration is a reason for moral judgment (indicated in Table 5 as 'Reason') or not (indicated in Table 5 as 'Not a Reason', except for Probability). The researchers found that, in general, when indicating that the specific consideration is a reason for moral judgment more participants showed preference for indirect harm and indicated the direct action as more wrong (ranging from 15% to 16.9%) whereas fewer participants indicate that the indirect action is more wrong (ranging from 2.8% to 3.9%). Similarly, when indicating that the specific consideration is *not* a reason for moral judgment, more participants showed preference for indirect harm and indicate the direct action as more wrong (ranging from 17.4% to 32.4%) whereas fewer participants indicated that the indirect action is more wrong (ranging from 5.1% to 6.5%).

In the replications we conducted, findings were broadly in line with the results of Royzman and Baron (2002). When indicating that the specific consideration is a reason for moral judgment, more participants showed preference for indirect harm and indicated that the direct action is more wrong (ranging from 17.6% to 66.7%) whereas fewer participants indicated that the indirect action is more wrong (ranging from 0% to 12.3%). Similarly, when indicating that the specific consideration is *not* a reason for moral judgment more participants showed a preference for indirect harm and indicate direct action is more wrong (ranging from 13.4% to 51.6%) whereas fewer participants indicated that the indirect action is more wrong (ranging from 6.5% to 29%). Overall, in all cases the proportions of participants who indicated the direct action is more wrong (thus indicating preference for indirect harm) were larger than the proportion of participants indicating the indirect action is more wrong, irrespective of whether they considered the specific consideration to be a reason for moral judgment or not.

Overall, in both replication attempts, we successfully replicated the correlational evidence that Royzman and Baron (2002) presented when investigating potential factors underlying the preference for indirect harm (probability of harm, intent, appearance, omission, and directness). This suggests that all of these factors likely play a part as psychological underpinnings of the preference for indirect harm. However, the evidence presented is correlational and only shows a statistical association rather than a neat cause-effect path. Further research may experimentally investigate the causality of these associations by manipulating intent or appearance within direct and indirect harm, for example. This is especially interesting in light of the literature showing that moral judgment in general, and preference for indirect harm in particular, are intuitive processes for which people are unable to quickly provide a justification (Cushman et al., 2006; Paharia et al., 2009) or explain why people prefer indirect harm to direct harm.

Mini meta-analysis effect summary

We summarized the findings of the two replications studies together with the target article original findings using mini meta-analyses for each of the scenarios to assess the overall effect size (Goh, Hall, & Rosenthal, 2016; Lakens & Etz, 2017 - see plots in Figure 2). The overall effects for the organ scenario was d = 0.47, CI [0.18, 0.75], and for the zoo scenario d = 0.46 [0.26, 0.65]. We conclude that the two scenarios had comparable weak to medium effects that are different from null.

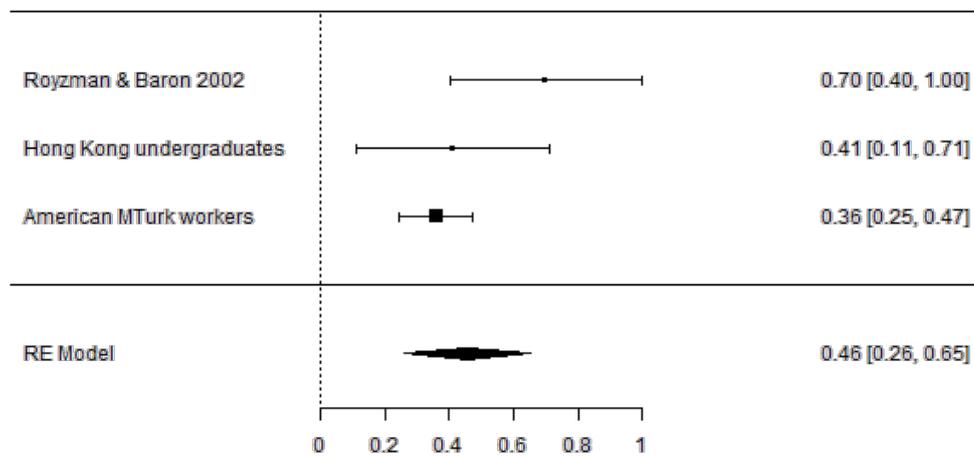## Organ donor scenario
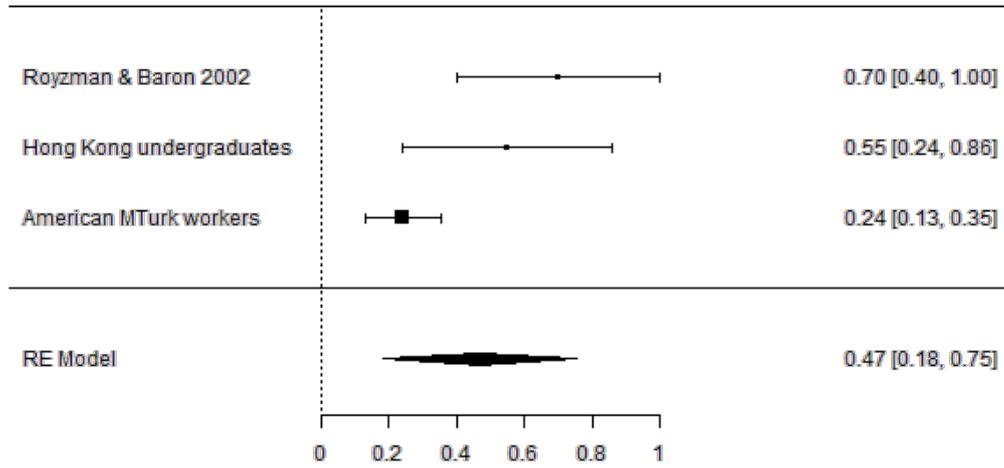
Zoo scenario

Figure 2. Mini meta-analysis effect size estimates (Cohen's d) and 95% confidence intervals (CIs) around effect size estimates for the original study and the two replication attempts in the two scenarios.

Table 5

Reasons for morality and the preference for indirect harm effect: frequencies and correlations

Hong Kong undergraduate Sample					American MTurk sample				
Organ donor		Predicted (Indirect)	Opposite (Direct)	Morality *r*	Organ donor		Predicted (Indirect)	Opposite (Direct)	Morality *r*
Probability		28.3%	0%	.144 [-0.15, 0.41]	Probability		18.8%	10.2%	.331 [0.23, 0.43]***
Directness					Directness				
	Reason	50.0%	7.1%	.631 [0.34, 0.81]***		Reason	29.4%	7.1%	.555 [0.45, 0.64]***
	Not a Reason	43.3%	16.7%	-.058 [-0.41, 0.31]		Not a Reason	32.9%	10.5%	.101 [-0.03, 0.23]
Appearance					Appearance				
	Reason	48.1%	0%	.291 [-0.10, 0.60]		Reason	24.4%	8.8%	.400 [0.28, 0.51]***
	Not a Reason	45.5%	12.1%	-.110 [-0.44, 0.24]		Not a Reason	27.3%	12.8%	.293 [0.17, 0.40]***
Omission					Omission				
	Reason	24.3%	5.4%	.714 [0.51, 0.84] ***		Reason	21.1%	7.7%	.424 [0.32, 0.52]***
	Not a Reason	17.1%	8.6%	.262 [-0.08, 0.55]		Not a Reason	18.5%	9.5%	.228 [0.11, 0.34]***
Intent					Intent				
	Reason	31.4%	2.9%	.685 [0.46, 0.83] ***		Reason	21.1%	5.1%	.363 [0.25, 0.47]***
	Not a Reason	17.6%	14.7%	.080 [-0.27, 0.41]		Not a Reason	17.0%	6.5%	.166 [0.04, 0.28]***
Zoo		Predicted (Indirect)	Opposite (Direct)	Morality *r*	Zoo		Predicted (Indirect)	Opposite (Direct)	Morality *r*
Probability		26.1%	15.2%	.499 [0.24, 0.69]***	Probability		13.4%	10.8%	.309 [0.21, 0.41]***
Directness					Directness				
	Reason	66.7%	4.8%	.532 [0.13, 0.78]***		Reason	47.3%	12.2%	.612 [0.52, 0.69]***
	Not a Reason	51.6%	29.0%	-.029 [-0.38, 0.33]		Not a Reason	43.2%	13.5%	.191 [0.05, 0.32]***
Appearance					Appearance				
	Reason	58.3%	4.2%	.894 [0.77, 0.95]***		Reason	38.7%	12.3%	.630 [0.54, 0.71]***
	Not a Reason	41.9%	29.0%	.191 [-0.18, 0.51]		Not a Reason	41.9%	10.5%	.290 [0.16, 0.41]***
Omission					Omission				
	Reason	28.1%	9.4%	.707 [0.48, 0.85]***		Reason	30.0%	10.9%	.485 [0.38, 0.58]***
	Not a Reason	29.4%	11.8%	.220 [-0.13, 0.52]		Not a Reason	28.2%	10.0%	.171 [0.04, 0.30]*
Intent					Intent				
	Reason	17.6%	11.8%	.324 [-0.02, 0.60]		Reason	31.7%	11.1%	.511 [0.41, 0.60]***
	Not a Reason	22.2%	11.1%	-.199 [-0.50, 0.14]		Not a Reason	24.3%	9.5%	.132 [0.00, 0.26]

Note. Correlation (Pearson's *r*) between morality and considerations, according to whether or not it was cited as a reason for a moral judgment. 'Predicted' indicates the share of responders indicating that the direct action was more wrong; 'Opposite' indicates the share of responders indicating that the indirect action was more wrong. * *p* < .05; ** *p* < .01; *** *p* < .001.

General Discussion

We successfully replicated findings from Royzman and Baron (2002) Study 2 with a non-naive undergraduate sample from Hong Kong and an American MTurk sample. These results provide empirical support for the preference for indirect harm phenomenon: people tend to prefer indirect harm over direct harm. We summarize the replications as "signal and consistent" according to the LeBel et al.'s (2019) replication success criteria, yet we note that equivalence tests indicated overall weaker effects compared to the target article findings. Mini meta-analyses of the replications and original findings indicated weak to medium effects that are different from null.

What may explain the weaker effects? Sample and time are the typical suspects. Royzman and Baron (2002) study was conducted using an Internet sample, resembling the MTurk sample in the replications, although MTurk workers are likely more experienced in participating in online studies (Chandler, Mueller, & Paolacci, 2014). Compared to the original sample, the Hong Kong sample was of a different cultural and linguistic background and had a much higher familiarity with heuristics and biases. We believe, however, that both sample and the passing of time are limited explanations given our other judgment and decision-making replications with similar samples showing high consistency between these two samples and the original findings (e.g., Chandrashekar et al, 2020; Chen et al., 2020). We cannot, however, rule out any possibility with confidence, and the many differences between the original study and our replications make it difficult to determine the cause. A possible future direction is to conduct a meta-analysis on the literature testing for moderators.

Our findings suggest that the classic phenomenon is replicable, yet that we may need to update our expectations regarding effect size. Replications are especially useful in this regard. Researchers can now use the replications' effect-size as an updated and more conservative estimate of the effect when designing their follow-up studies.

Author Contact

Ignazio Ziano, Ignazio.ZIANO@grenoble-em.com, orcid.org/0000-0002-4957-3614
Yu Jie Wang, u3529917@connect.hku.hk,
Sydney Susanto Sany, sydneyssany@yahoo.com
Long Ho Ngai, sngai717@connect.hku.hk
Yuk Kwan Lau, tonilau@connect.hku.hk
Iban Kaur Bhattal, iban03@connect.hku.hk
Pui Sin Keung, u3534402@connect.hku.hk
Yan To Wong, norawyt@connect.hku.hk
Wing Zhang Tong, u3544235@connect.hku.hk
Bo Ley Cheng, boleystudies@gmail.com
Hill Yan Cedar Chan, cedar@hku.hk
Gilad Feldman (corresponding author), gfeldman@hku.hk,
orcid.org/0000-0003-2812-6599

Author Contributions

Gilad Feldman (corresponding author - GF from now on and in the table below) was the course instructor for fundamentals and advanced social psychology courses (PSYC2020/3052) and led the two reported replication efforts in those courses. GF supervised each step in the project, conducted the pre-registrations, and ran data collection. Ignazio Ziano (joint first author- IZ from now on and in the table below) integrated the two replication efforts into a manuscript with validation and further extensions of the statistical analyses. GF and IZ jointly finalized the manuscript for submission.

Yu Jie Wang and Sydney Susanto Sany (joint first authors) conducted the US replication as part of the advanced social psychology course (identified as Students PSYC3052 in the table below).

Long Ho Ngai, Yuk Kwan Lau, Iban Kaur Bhattal, Pui Sin Keung, Yan To Wong, and Wing Zhang Tong conducted the Hong Kong replication as part of the fundamentals of social psychology course (joint fourth authors; identified as Students PSYC2020 in the table below).

Bo Ley Cheng (Teaching Assistant; included in the "TAs" column in the table below) guided and assisted the replication effort in the PSYC3052 course.

Hill Yan Cedar Chan guided and assisted the replication effort in the PSYC2020 course (Teaching Assistant; included in the "TAs" column in the table below).

Contributor Roles Taxonomy

In the table below, employ CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to the url (https://www.casrai.org/credit.html) on details and definitions of each of the roles listed below

Role	IZ	GF	Students PSYC 2020	Students PSYC 3052	TAs
Conceptualization		X			
Pre-registrations		X	X	X	
Data curation		X			
Formal analysis	X	X	X	X	
Funding acquisition		X			
Investigation		X	X	X	
Methodology		X		X	
Pre-registration peer review / verification	X		X	X	X
Data analysis peer review / verification	X		X	X	
Project administration		X			X
Resources		X			
Software	X	X	X	X	
Supervision					X
Validation	X	X			
Visualization	X				
Writing-original draft	X	X			
Writing-review and editing	X	X			

Open Science Practices

This article earned the Preregistration+, Open Data and the Open Materials badge for preregistering the hypothesis and analysis before data collection, and for making the data and materials openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

Bennett, J. (1966). 'Whatever the Consequences'. *Analysis*, 26, 83–102.

Chandrashekar, S. P., Yeung., S., Yau, K., Feldman, G., … (2020). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and extensions linking to free will beliefs. DOI: 10.13140/RG.2.2.19878.16961 Manuscript under review. Retrieved December 2019 from https://www.researchgate.net/publication/331431431_Agency_and_self-other_asymmetries_in_perceived_bias_and_shortcomings_Replications_of_the_Bias_Blind_Spot_and_extensions_linking_to_free_will_beliefs

Chen, J., Hui, L. S., Yu, T., Feldman, G., Zeng, S. V., Ching, T. L., … & Cheng, B. L. (2020). Foregone opportunities and choosing not to act: Replications of Inaction Inertia effect. Social Psychological and Personality Science. Manuscript accepted for publication. Retrieved December 2019 from: https://www.researchgate.net/publication/332550110_Foregone_opportunities_and_choosing_not_to_act_Replications_of_Inaction_Inertia_effect

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. Psychological Science, 17, 1082-1089.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. Social and Personality Psychology Compass, 10, 535-549.

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals 1. Journal of Applied Social Psychology, 31, 191-221.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. Mind & Language, 22, 1-21.

Jamison, J., Yay, T., & Feldman, G. (2020). Action-in-action asymmetries in moral scenarios: Replication of the omission bias examining morality and blame with extensions linking to causality, intent, and regret. Manuscript under review. Retrieved December 2019 from https://www.researchgate.net/publication/326260685_Action-inaction_asymmetries_in_moral_scenarios_Replication_of_the_omission_bias_examining_morality_and_blame_with_extensions_linking_to_causality_intent_and_regret

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and non-significant findings are probably true. Social Psychological and Personality Science, 8, 875-881.

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, 259-269.

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, 389-402.

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A Brief Guide to Evaluate Replications. Meta Psychology, 541, 1–17. https://doi.org/10.31219/osf.io/paxyn

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. Behavior Research Methods, 49, 433-442.

Milgram, S. (1974). Obedience to authority. An Experimental View. New York: Harper.

Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. Organizational Behavior and Human Decision Processes, 109, 134-141.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science, 23, 184-188.

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. Social Justice Research, 15, 165-184.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. Journal of Experimental Social Psychology, 27, 7