Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–8

DOI: 10.1177/0956797620948821
www.psychologicalscience.org/PS


Moral Choice When Harming Is Unavoidable

Jonathan Z. Berman[1] and Daniella Kupor[2]
[1]Department of Marketing, London Business School, and [2]Department of Marketing, Boston University

Abstract

Past research suggests that actors often seek to minimize harm at the cost of maximizing social welfare. However, this prior research has confounded a desire to minimize the negative impact caused by one's actions (harm aversion) with a desire to avoid causing any harm whatsoever (harm avoidance). Across six studies (N = 2,152), we demonstrate that these two motives are distinct. When decision-makers can completely avoid committing a harmful act, they strongly prefer to do so. However, harming cannot always be avoided. Often, decision-makers must choose between committing less harm for less benefit and committing more harm for more benefit. In these cases, harm aversion diminishes substantially, and decision-makers become increasingly willing to commit greater harm to obtain greater benefits. Thus, value trade-offs that decision-makers refuse to accept when it is possible to completely avoid committing harm can suddenly become desirable when some harm must be committed.

Keywords

moral choice, value trade-offs, harm aversion, harm avoidance, protected values, open data, open materials, preregistered

Received 12/20/19; Revision accepted 5/15/20

To what extent are individuals harm averse? Existing research suggests that harm aversion is a principle force guiding moral judgment and choice. Individuals are reluctant to harm others even when doing so would maximize social welfare (Petrinovich, O'Neill, & Jorgensen, 1993) and even prefer to harm themselves rather than harm others (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014).

We argue that harm aversion is more limited in scope than previously documented. Consistent with prior literature, our hypothesis was that individuals strongly prefer to avoid committing a harmful act if they are able to do so (Miller, Hannikainen, & Cushman, 2014; Petrinovich et al., 1993; Spranca, Minsk, & Baron, 1991). However, harming is not always avoidable, and decision-makers must often choose between committing small amounts of harm to obtain few social benefits and committing larger amounts of harm to obtain greater social benefits. We argue that in these situations, individuals become less likely to prefer that harm be minimized and more likely to prefer that social benefits be maximized. Thus, although individuals are reluctant to trade off *no harm* for *some harm* to achieve benefits, they are much more willing to trade off *some harm* for *more harm* in return for the same degree of—or even fewer—marginal benefits.

Choices that necessitate the commission of at least some harm pervade decision-making. Most everyday consumption decisions require people to choose among alternatives that harm the environment to varying degrees (Csikszentmihalyi, 2000). For example, when deciding where to live or how to furnish one's home, individuals often must choose among options that each carry positive carbon footprints. Similarly, policymakers often must choose among options that each furnish a mixture of societal harms and benefits of varying magnitudes. Yet we know little about how individuals balance the costs and benefits of their actions when all the options in a choice set require an

Corresponding Author:
Jonathan Z. Berman, London Business School, Sussex Place, Regent's Park, London NW1 4SA, United Kingdom
E-mail: jberman@london.edu

actor to commit some harm (cf. Bonnefon, Shariff, & Rahwan, 2016).

Distinguishing Between Dilemmas Containing Avoidable and Unavoidable Harmful Acts

We distinguish between dilemmas in which decision-makers can avoid committing a harmful act and dilemmas in which committing a harmful act is unavoidable. In the former type of dilemma, individuals can select an option that does not require the commission of harm. Consider a government agent faced with two proposals for feeding the hungry: Proposal A would create enough farmland to feed 100 families without harming the rain forest, whereas proposal B would create enough farmland to feed 500 families but would require razing one acre of the rain forest. If the agent chooses proposal A, he or she would avoid committing a harmful act (destroying the rain forest) but would also fail to maximize benefits (feeding as many people as possible).

In the latter type of dilemma, decision-makers must instead choose between options that would each require a harmful act to be committed. Imagine now that the two proposals would feed the same number of families, but proposal A would require razing one acre of the rain forest, whereas proposal B would require razing two acres. Across both dilemmas, the marginal increase in harms and benefits is matched: Destroying one additional acre of the rain forest would provide food for an additional 400 families. If an agent refuses to destroy one additional acre in the former case, the agent ought to be similarly unwilling to do so in the latter case. However, when we randomly assigned each of 380 online participants to one of these two scenarios, 46.1% chose to destroy one additional rain-forest acre in the former case, when it was possible to completely avoid committing a harmful act, whereas 78.4% chose to do so in the latter case, when completely avoiding committing harm was no longer possible, $\chi^2(1, N = 380) = 41.53$, $p < .001$, $\varphi = .33$ (this study was preregistered at https://aspredicted.org/4gf64.pdf). This suggests that value trade-offs that individuals refuse to accept when it is possible to completely avoid committing harm can suddenly become desirable when some harm must be committed.

Why might harm aversion diminish when individuals can no longer completely avoid committing a harmful act? Descriptive theories of morality suggest that judgments often reflect an adherence to intuitively appealing deontological principles such as "destroying the rain forest is wrong, regardless of the consequences" (e.g., Everett, Pizarro, & Crockett, 2016). We argue that the intuitive appeal of many deontological principles breaks down when decision-makers cannot altogether avoid committing harm. In these situations, individuals can no longer completely avoid violating a principle because the principle will be violated regardless of what is chosen. Decision-makers may thus become concerned with making sure that the violation is not done in vain. Instead they may prefer that more harm be conducted to ensure that the violation of a principle is "worth it," similar to how people are willing to spend additional resources to justify sunk costs (Arkes & Blumer, 1985) or recuperate losses (Tversky & Kahneman, 1991).

Overview

In the following studies, we examined how individuals make value trade-offs across dilemmas containing avoidable and unavoidable harmful acts. We predicted that the preference to minimize harm would decrease when decision-makers were no longer able to completely avoid committing a harmful act and instead had to choose between options that each require committing some harm.

In all studies, sample sizes were determined in advance. All conditions and measures assessed are reported in this article with the exception of attention-check questions, which can be found in the supporting materials at https://osf.io/4vd53/, alongside full study materials. In all studies, we recruited a minimum of 100 participants per cell. Data files for all studies can be accessed on OSF at https://osf.io/84yhe/.

Study 1: Pulling Life Support From the Terminally Ill

Study 1 presents further evidence that people are resistant to exchanging harm for benefits when they can completely avoid committing a harmful act but are more likely to do so when committing some harm is unavoidable.

Method

We recruited 405 participants from Amazon Mechanical Turk (MTurk) to participate in this study. We removed 34 participants who failed an attention check, resulting in a final sample of 371 participants (mean age = 37.3 years; 55% female, 45% male).

Each participant was randomly assigned to either an avoidable-harm condition or an unavoidable-harm condition. All participants assumed the role of a doctor who faced a decision regarding whether to pull life support from a dying patient to save money that would be used for cancer research. In the avoidable-harm condition, participants could choose to pull life support from one dying child to save $60,000 for cancer research

or choose not to do so. In the unavoidable-harm condition, participants instead faced a decision about whether to pull life support from one dying child to save $800 for cancer research or to pull life support from two dying children to save $50,000 for cancer research. Despite the fact that the termination of one additional life would yield greater benefits in the avoidable-harm condition (a net savings of $60,000) than in the unavoidable-harm condition (a net savings of $49,200), we predicted that participants in the avoidable-harm condition would be less likely to accept such a trade-off than those in the unavoidable-harm condition.

Results

Consistent with our hypothesis, results showed that participants were more likely to prefer to commit greater harm in exchange for greater benefits in the unavoidable-harm condition (69.9%) than in the avoidable-harm condition (28.6%), $\chi^2(1, N = 371) = 65.10$, $p < .001$, $\varphi = .41$.[1]

Study 2: Donating to the Opposition

In Study 2, we examined whether our previous results would hold in an incentive-compatible context—allocating money to support political nonprofits. We predicted that when participants could avoid allocating any money to support an oppositional political party, they would do so, even if doing so failed to maximize net benefits to support their own party. However, we predicted that when participants could no longer completely avoid allocating at least some money to the opposition, they would become more likely to maximize net benefits to their own party.

Method

We preregistered the design, sample size, and analysis plan for this study on AsPredicted (https://aspredicted.org/8a6ta.pdf). We aimed to recruit 400 participants from MTurk and collected a total of 403 participants. Following our preregistration plan, we removed 72 participants who failed an attention check, resulting in a final sample of 331 participants (mean age = 38 years; 47.4% male, 52.0% female, 0.6% nonbinary).

Following our preregistration plan, we restricted participation to individuals who were registered with either the Democratic (57.1%) or Republican (42.9%) party. Participants were faced with a choice regarding the allocation of real money to two nonprofits. One of these nonprofits (Priorities USA) was dedicated to propelling the Democratic presidential nominee to victory, whereas the other nonprofit (WinRed) was dedicated to propelling the Republican presidential nominee to victory. We operationalized benefits as a donation to a nonprofit in support of participants' registered political party and operationalized harm as a donation in support of the opposing party. A pretest confirmed that participants perceived these two types of donations as yielding benefits and harms, respectively (for details, see the Results section).

We randomly assigned each participant to either an avoidable-harm condition or an unavoidable-harm condition. In the avoidable-harm condition, participants could choose to donate $2 to the nonprofit supporting their own political party and $0 to the nonprofit supporting the political opposition (a harm-minimization option), or they could choose to donate $4 to a nonprofit supporting their own political party and $1 to a nonprofit supporting the political opposition (a benefit-maximization option). For example, participants who indicated that they were registered Democrats could choose to donate either (a) $2 to Priorities USA and $0 to WinRed or (b) $4 to Priorities USA and $1 to WinRed.

In the unavoidable-harm condition, the options were the same except that we added an additional $2 donation to the political-opposition nonprofit; as a result, participants could no longer avoid committing a harmful act (i.e., donating to the opposing party). Thus, the harm-minimization option would result in a $2 donation in support of the participant's own party and a $2 donation in support of the political opposition, whereas the benefit-maximization option would result in a $4 donation in support of the participant's own party and a $3 donation in support of the political opposition. For example, participants who indicated that they were registered Democrats could choose to donate either (a) $2 to Priorities USA and $2 to WinRed or (b) $4 to Priorities USA and $3 to WinRed.

In sum, across both conditions, choosing to minimize harm meant that participants were unwilling to donate an additional $1 to the political opposition in order for their own party to receive an additional $2, whereas choosing to maximize net benefits meant that they were willing to make this trade-off.

Finally, before making their selection, all participants learned that in addition to their participation payment, they would be entered into a raffle. If they won the raffle, a donation would be made on their behalf in accordance with their selection in the survey.

Results

Pretest. We conducted a pretest among 78 participants drawn from MTurk to assess whether participants perceived the donation of funds to support their own political party as a social benefit and the donation of funds to support

the opposing party as a social harm. Participants rated the extent to which a donation made to each nonprofit was either beneficial or harmful to the country on a 7-point bipolar scale (−3 = *extremely harmful*, 0 = *neither beneficial nor harmful*, 3 = *extremely beneficial*). Participants judged a donation made to the nonprofit aligned with their own party to be more beneficial than harmful ($M = 1.24$, $SD = 1.38$), $t(77) = 7.97$, $p < .001$, $d = 0.90$, 95% confidence interval (CI) = [0.60, 1.20], whereas they judged a donation made to the nonprofit aligned with the opposing party to be more harmful than beneficial ($M = -0.82$, $SD = 1.81$), $t(77) = -3.99$, $p < .001$, $d = 0.45$, 95% CI = [0.21, 0.70].

Main study. Consistent with our preregistered hypothesis, results showed that participants were more likely to trade off greater harm for greater benefits in the unavoidable-harm condition (67.5%) than in the avoidable-harm condition (35.1%), $\chi^2(1, N = 331) = 34.68$, $p < .001$, $\varphi = .32$. In other words, when it was possible to completely avoid committing a harmful act (donating to support the opposition), participants strongly preferred to do so even though it deprived their own political party of even more money. However, when it was not possible to completely avoid doing so, participants preferred to maximize net benefits to their own party rather than minimize contributions to the opposition.

Discussion

Study 2 found that participants favored minimizing harm when committing a harmful act was avoidable, whereas they favored maximizing benefits when committing at least some harm was unavoidable.

Study 3: Divergent Goals Activated by Avoidable and Unavoidable Harmful Acts

Study 3 assessed people's reasoning for their choices. We predicted that individuals rely more heavily on a harm-minimization goal when they can completely avoid committing a harmful act but rely more heavily on a benefit-maximization goal when completely avoiding a harmful act is impossible.

Method

We preregistered the design, sample size, and analysis plan for this study on AsPredicted (https://aspredicted.org/vp4rg.pdf). We aimed to recruit 200 participants from MTurk and collected a total of 203 participants. Following our preregistration plan, we removed 10 participants who failed an attention check, resulting in a final sample of 193 participants (mean age = 40 years; 67% female, 33% male).

Each participant was again randomly assigned to either an avoidable-harm condition or an unavoidable-harm condition. As in the previous studies, participants in both conditions chose between two options—a harm-minimizing option or a benefit-maximizing option. In the avoidable-harm condition, the harm-minimizing option involved using a type of harmless fertilizer that would enable the growth of additional crops sufficient to provide food to 1 million hungry children. In contrast, the benefit-maximizing option involved using an alternate fertilizer that would enable the growth of additional crops sufficient to provide food for 4 million hungry children but would cause a 0.5-in. hole in the ozone layer. In the unavoidable-harm condition, participants were faced with the same trade-off between whether or not a 0.5-in. hole in the ozone layer should be created to supply food for 3 million children, but this time, both options caused an additional 1-in. hole in the ozone layer. Specifically, the harm-minimizing option enabled the growth of enough food for 1 million children at the cost of creating a 1-in. hole in the ozone, whereas the benefit-maximizing option enabled the growth of enough food for 4 million children at the cost of creating a 1.5-in. hole in the ozone. In sum, across both conditions, participants faced the decision of whether a 0.5-in. hole in the ozone layer should be created to provide food for 3 million hungry children.

After making their decision, participants completed two additional measures to examine the reason underlying their decision. The first question captured the impact of a harm-minimization motive on participants' decisions by asking them to indicate the extent to which their decision was driven by a desire to do as little harm as possible to the ozone layer. The second question captured the impact of a benefit-maximization motive on participants' decisions by asking them to indicate the extent to which their decision was driven by a desire to ensure that any harm to the ozone layer provided sufficient benefit. Participants indicated their responses to these two measures on separate 7-point scales (1 = *not at all*, 7 = *very much*).

Results

Consistent with our preregistered hypothesis, results showed that participants were less likely to commit greater harm in exchange for greater benefits in the avoidable-harm condition (21.6%) than in the unavoidable-harm condition (61.5%), $\chi^2(1, N = 193) = 31.91$, $p < .001$, $\varphi = .41$.

In accordance with our preregistered analysis plan, we computed a difference score by subtracting participants' responses to the benefit-maximization measure from their responses to the harm-reduction measure to

examine the relative impact of these two goals on participants' decisions. Thus, a positive difference score reflected a greater reliance on a harm-reduction goal as opposed to a benefit-maximization goal. As predicted, analysis of this difference score indicated that the harm-reduction objective more strongly guided decision-making than the benefit-maximization objective in the avoidable-harm condition (M = 1.60, SD = 3.32) than in the unavoidable-harm condition (M = −1.13, SD = 2.94), $t(192) = 6.03$, $p < .001$, $d = 0.87$, 95% CI = [0.57, 1.16].

Also as expected, a bootstrap mediation analysis with 10,000 samples (in which condition was entered as the independent variable, harm decisions were entered as the dependent variable, and the difference score was entered as the mediator) revealed a significant indirect effect (95% CI = [0.19, 0.36]). The harm-reduction goal more strongly guided decision-making than the benefit-maximization goal in the avoidable-harm condition than in the unavoidable-harm condition (path *a:* $b = 2.75$, $SE = 0.46$, $p < .001$), and the more strongly the harm-reduction goal (relative to the benefit-maximization goal) guided decision-making, the more likely participants were to minimize harm (path *b:* $b = 0.10$, $SE = 0.01$, $p < .001$). When the mediator was controlled for, the effect of condition on harm decisions was significantly reduced (path *c:* $b = 0.40$, $SE = 0.07$, $p < .001$; path c': $b = 0.13$, $SE = 0.05$, $p = .013$).

Study 4: Increasing Versus Diminishing Marginal Returns to Harm

In Study 4, we examined the robustness of our findings. In our previous studies, the options in the unavoidable-harm conditions furnished increasing marginal returns to committing greater harm (e.g., pulling life support from a second child provided a greater marginal increase in money saved than pulling life support from the first). It is therefore possible that individuals are willing to exchange greater harm for greater benefit only when doing so produces efficiency gains (cf. de Langhe & Puntoni, 2014). If so, we would expect participants to be particularly averse to committing greater harm when it furnishes diminishing marginal returns (e.g., pulling life support from a second child provides a smaller marginal increase in money saved than pulling life support from the first). However, we propose that the desire to completely avoid committing a harmful act holds a particularly strong influence on choice. Thus, we predicted that decision-makers would be more likely to minimize harm when it was possible to avoid committing any harm relative to when choice sets offered diminishing marginal returns to committing greater harm.

Method

We preregistered the design, sample size, and analysis plan for this study on AsPredicted (https://aspredicted.org/3kq9y.pdf). We aimed to recruit 500 participants from MTurk and obtained a total sample of 509 participants. Following our preregistration plan, we removed 28 participants who failed an attention check, resulting in a final sample of 481 participants (mean age = 37 years; 55% female, 45% male).

All participants read an adapted version of the classic trolley problem (Foot, 1967; Thomson, 1976). In this version, a trolley without passengers and without a conductor was traveling at full speed down a track toward 11 people who would all die if nothing changed. In the avoidable-harm condition, participants faced a decision about whether to push an individual onto the tracks, a decision that would kill this individual but would slow the trolley enough to save three people. Whereas choosing to push the person would maximize benefits, it also would require committing a harmful act.

In the diminishing-marginal-returns-to-harm condition, participants faced a decision about whether to push one individual onto the tracks in order to slow the trolley enough to save three people or to push two individuals onto the tracks in order to slow the trolley enough to save five people. Thus, both options required committing harmful acts. Moreover, the marginal benefits achieved from pushing a second person (two lives saved) were smaller than the benefits achieved from pushing the first person onto the tracks (three lives saved).

Finally, in the increasing-marginal-returns-to-harm condition, participants faced a decision about whether to push one individual onto the tracks in order to slow the trolley enough to save two people or to push two individuals onto the tracks in order to slow the trolley enough to save five people. Again, both options required committing harmful acts. Moreover, the marginal benefits achieved from pushing a second person (three lives saved) were larger than the benefits achieved from pushing the first person onto the tracks (two lives saved).

The decision that participants faced in the present study equated harms and benefits on the same scale (lives sacrificed vs. lives saved) and therefore allowed for a clear way to calculate the expected value of outcomes. In particular, in all conditions, the benefit-maximizing option led to a larger number of net lives saved than the harm-minimizing option. Thus, if actors become more consequentialist when the option to completely avoid committing any harm is removed, then they would be more likely to choose the benefit-maximization option (i.e., the option that maximizes social welfare) in both

the increasing-marginal-returns-to-harm and diminishing-marginal-returns-to-harm conditions than in the avoidable-harm condition.

Results

Consistent with our preregistered hypothesis, participants were less likely to commit greater harm for greater benefits in the avoidable-harm condition (25.5%) than in both the diminishing-marginal-returns-to-harm condition (38.6%), $\chi^2(1, N = 319) = 6.33$, $p = .01$, $\varphi = .14$, and the increasing-marginal-returns-to-harm condition (59.3%), $\chi^2(1, N = 364) = 37.75$, $p < .001$, $\varphi = .34$. Additionally, participants were less likely to commit greater harm for greater benefits in the diminishing-marginal-returns-to-harm condition than in the increasing-marginal-returns-to-harm condition, $\chi^2(1, N = 320) = 13.65$, $p < .001$, $\varphi = .21$.

Discussion

Although participants were more likely to commit more harm when it furnished increasing marginal returns, a preference for increasing marginal returns to harm cannot fully explain our findings. Moreover, it appears that actors become more consequentialist when committing harm is unavoidable.[2]

Study 4 further shows that it is the unavoidability of harmful *acts*—not the unavoidability of harmful *outcomes*—that increases willingness to exchange greater harms for greater benefits. Across all conditions, a harmful outcome was unavoidable—at least one person would die regardless of what was chosen. However, it was only when participants could altogether escape committing a harmful act (in the avoidable-harm condition) that they were particularly resistant to trading off greater harm for greater benefits.

Study 5: Protected Values Become Corrupted When Committing Harm Is Unavoidable

Protected values are those that people consider infinitely important and ought not be traded off under any circumstance (Baron & Spranca, 1997; Tetlock, Kristel, Elson, Green, & Lerner, 2000). In Study 5, we examined whether people who believe that harming violates a protected value would be particularly reluctant to exchange greater harm for greater benefits even when some harm must be committed.

On the one hand, proclaiming that a harm violates an infinitely important protected value may cause individuals to minimize that harm irrespective of the decision context. On the other hand, some individuals are willing to put a price on protected values when pressed (Baron & Leshner, 2000). Protected values may thus better reflect a resistance to engage in a negotiation over a value rather than an absolute resistance to making value trade-offs per se. These individuals may further find it particularly tragic to commit harm without sufficient benefit.

We expected that people are particularly averse to making value trade-offs that violate a protected value when it is possible to avoid committing any violation of it. However, we also expected that when individuals must commit some harm that violates a protected value, they will become more willing to exchange greater harm for greater benefit. In contrast, we expected that people who do not believe that a value is sacred—and thus do not object to making value trade-offs—would display greater preference consistency across both types of dilemmas.

Method

We preregistered the design, sample size, and analysis plan for this study on AsPredicted (https://aspredicted.org/rz98j.pdf). We aimed to recruit 400 participants from MTurk and collected a total of 410 participants. Following our preregistration plan, we removed 14 participants who failed an attention check, resulting in a final sample of 396 participants (mean age = 37 years; 49% female, 51% male).

All participants first indicated whether they perceived preserving tropical rain-forest trees to be a protected value. In particular, participants indicated which one of three options best matched their reaction to cutting down tropical rain-forest trees: "This is acceptable if it leads to some sort of benefits (money or something else) that are great enough," "This is not acceptable no matter how great the benefits," and "I do not object to this." Participants selected the button next to their preferred response. Following prior literature (Ritov & Baron, 1999), we coded only participants who selected the second response (76.2% of participants) as perceiving this harm as a violation of a protected value.

Next, all participants faced two decisions that imperiled the preservation of a tropical rain forest. Specifically, in the avoidable-harm condition, participants chose between two proposals: One proposed to build one water-treatment facility and would not require the destruction of any tropical rain-forest trees (i.e., the harm-minimizing option), and the other proposed to build five water-treatment facilities but required the destruction of one acre of tropical rain-forest trees (i.e., the benefit-maximizing option). In the unavoidable-harm condition, participants chose between a proposal to build one water-treatment facility that required the destruction of one acre of tropical rain-forest trees (i.e.,

the harm-minimizing option) and another proposal to build five water-treatment facilities that required the destruction of two acres of tropical rain-forest trees (i.e., the benefit-maximizing option). In sum, across both dilemmas, participants faced the decision of whether one acre of rain forest should be razed to enable the creation of four additional water-treatment facilities.

Participants viewed both dilemmas in a counterbalanced order. Thus, this study employed a 2 (protected value: yes, no) × 2 (dilemma: avoidable harm, unavoidable harm) mixed factorial design in which protected values constituted a measured between-participants factor and dilemma constituted a manipulated within-participants factor.

Results

We conducted a binary logistic mixed-effects regression that included random intercepts to control for within-participants variability and that regressed the decision context, self-reported protected values, and the interaction of these terms on participants' decisions. This regression revealed a significant interaction ($b = 1.20$, $SE = 0.51$, $z = 2.34$, $p = .019$), consistent with our preregistered hypothesis (see Fig. 1). Planned contrasts revealed that participants who perceived that the destruction of the rain forest violated a protected value were more likely to decide that one acre of this rain forest should be destroyed to enable the creation of four additional water-treatment facilities in the unavoidable-harm scenario (67.1%) than in the avoidable-harm scenario (42.0%; $b = 1.71$, $SE = 0.25$, $z = 6.77$, $p < .001$). By contrast, participants who did not perceive that the destruction of the rain forest violated a protected value were equally likely to decide that one acre of rain forest should be razed to enable the creation of four additional water-treatment facilities in both the unavoidable-harm scenario (74.0%) and the avoidable-harm scenario (68.4%; $b = 0.51$, $SE = 0.43$, $z = 1.22$, $p = .22$).

Participants who indicated that rain-forest preservation is sacred sought to minimize rain-forest destruction when it was possible to avoid committing any rain-forest destruction. However, when this was not possible, these individuals became much more willing to commit greater destruction to maximize benefits. Whereas past researchers have argued that individuals holding protected values are scope insensitive (Ritov & Baron, 1999), we found that this is not the case when the commission of harm is unavoidable.[3]

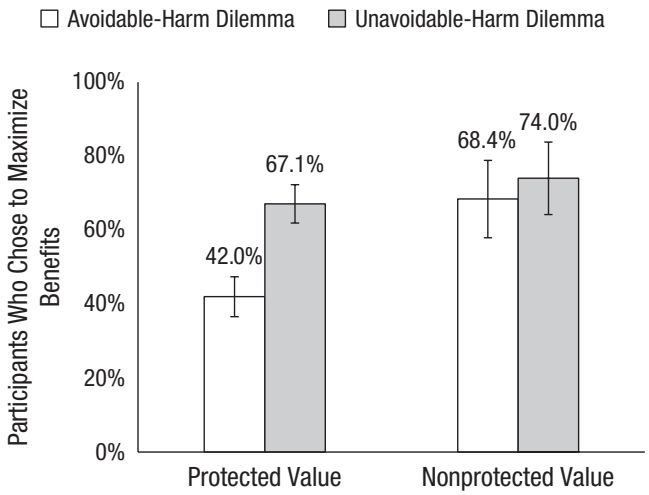

Fig. 1. Results from Study 5: percentage of participants who chose to commit greater harm in exchange for greater benefits by whether or not the participant indicated that rain-forest preservation is a protected value and type of ethical dilemma. Error bars correspond to 95% confidence intervals.

General Discussion

Across six studies, we demonstrated that the preference to avoid inflicting any harm not only is distinct from but also outweighs the preference to minimize its impact. Our results suggest that the manner in which individuals bracket instances of harm affects their willingness to commit harm (cf. Read, Loewenstein, Rabin, Keren, & Laibson, 1999). For instance, individuals may be more reluctant to commit a second violation a month after a first violation than they would be if the second violation occurred just moments after the first. This is because the two harmful actions may be more likely to be bracketed together in the latter case and may thus be perceived as an unavoidable-harm context.

Although we focused our examination on decisions impacting social welfare, similar outcomes may occur for decisions that are exclusively self-relevant. For instance, research suggests that individuals are particularly averse to holding debt if they do not need to be in debt but prefer to take on more debt to maintain their assets if holding debt is unavoidable (Sussman & Shafir, 2012).

Finally, in Study 4, we found that even when greater harm produced diminishing marginal benefits, individuals were still more willing to commit greater harm than when it was possible to commit no harm. However, there is likely a threshold for which committing more harm is no longer perceived as worthwhile. Future research can investigate factors that affect this threshold.

In sum, we found that decision-makers who can completely avoid committing a harmful act frequently choose to do so. However, when committing some harm is unavoidable, decision-makers become increasingly willing to trade off greater harm for greater benefits.

Transparency

Action Editor: Timothy J. Pleskac
Editor: D. Stephen Lindsay
Author Contributions

J. Z. Berman and D. Kupor jointly developed the concept and design for each study. J. Z. Berman collected and analyzed the data for Studies 1, 2, 4, 5, S1, and S2, and D. Kupor collected and analyzed the data for the introductory study and Studies 3, S3, and S4. Both authors jointly drafted the manuscript and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

All data and materials have been made publicly available via OSF and can be accessed at https://osf.io/84yhe/. The design and analysis plans for all but Study 1 were preregistered—introductory study: https://aspredicted.org/4gf64.pdf, Study 2: https://aspredicted.org/8a6ta.pdf, Study 3: https://aspredicted.org/vp4rg.pdf, Study 4: https://aspredicted.org/3kq9y.pdf, Study 5: https://aspredicted.org/rz98j.pdf, Study S1: https://aspredicted.org/yi5ps.pdf, and Study S2: https://aspredicted.org/nu5yt.pdf. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620948821. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Jonathan Z. Berman https://orcid.org/0000-0002-7214-0649

Notes

1. Study S1 (reported at https://osf.io/4vd53/) ruled out the possibility that participants were averse to harming a single child in the avoidable-harm condition because of the identifiable-victim effect (Kogut & Ritov, 2011).
2. For conceptual replications of Study 4, see Studies S2 and S3 at https://osf.io/4vd53/.
3. The within-participants design of Study 5 additionally ruled out a value-uncertainty explanation for our findings (see Study S4 at https://osf.io/4vd53/).

References

Arkes, H., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*, 124–140.

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied*, *6*, 183–194.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*, 1573–1576.

Crockett, M., Kurth-Nelson, Z., Siegel, J., Dayan, P., & Dolan, R. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences, USA*, *111*, 17320–17325.

Csikszentmihalyi, M. (2000). The costs and benefits of consuming. *Journal of Consumer Research*, *27*, 267–272.

de Langhe, B., & Puntoni, S. (2014). Bang for the buck: Gain-loss ratio as a driver of judgment and choice. *Management Science*, *61*, 1137–1163.

Everett, J., Pizarro, D., & Crockett, M. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*, 772–787.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5–15.

Kogut, T., & Ritov, I. (2011). The identifiable victim effect: Causes and boundary conditions. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The science of giving: Experimental approaches to the study of charity* (pp. 133–148). New York, NY: Psychology Press.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, *14*, 573–587.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467–478.

Read, D., Loewenstein, G., Rabin, M., Keren, G., & Laibson, D. (1999). Choice bracketing. In B. Fischhoff & C. F. Manski (Eds.), *Elicitation of preferences* (pp. 171–202). Dordrecht, The Netherlands: Springer.

Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, *79*, 79–94.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*, 76–105.

Sussman, A. B., & Shafir, E. (2012). On assets and debt in the psychology of perceived wealth. *Psychological Science*, *23*, 101–108.

Tetlock, P., Kristel, O., Elson, S., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853–870.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–217.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*, 1039–1061.