American Psychological Association

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

2020, Vol. 118, No. 2, 242–253
http://dx.doi.org/10.1037/pspa0000175

Being "Good" or "Good Enough": Prosocial Risk and the Structure of Moral Self-Regard

Julian J. Zlatev
Harvard University

Daniella M. Kupor
Boston University

Kristin Laurin
University of British Columbia

Dale T. Miller
Stanford University

The motivation to feel moral powerfully guides people's prosocial behavior. We propose that people's efforts to preserve their moral self-regard conform to a *moral threshold model*. This model predicts that people are primarily concerned with whether their prosocial behavior legitimates the claim that they have acted morally, a claim that often diverges from whether their behavior is in the best interests of the recipient. Specifically, it predicts that for people to feel moral following a prosocial decision, that decision need not have promised the greatest benefit for the recipient but only one larger than at least one other available outcome. Moreover, this model predicts that once people produce a benefit that exceeds this threshold, their moral self-regard is relatively insensitive to the magnitude of benefit that they produce. In 6 studies, we test this moral threshold model by examining people's prosocial risk decisions. We find that, compared with risky egoistic decisions, people systematically avoid making risky prosocial decisions that carry the possibility of producing the worst possible outcome in a choice set—even when this means avoiding a decision that is objectively superior. We further find that this aversion to producing the worst possible prosocial outcome leads people's prosocial (vs. egoistic) risk decisions to be less sensitive to those decisions' maximum possible benefit. We highlight theoretical and practical implications of these findings, including the detrimental consequence that people's desire to protect their moral self-regard can have on the amount of good that they produce.

Keywords: moral self-regard, morality, prosocial decisions, risk

The nature of morality has concentrated the minds of philosophers for millennia, and more recently has drawn the attention of psychologists and biologists (Batson, 2016; Bloom, 2012; de Waal, 2009; Greene, 2013; Hamlin, 2013). For many contemporary moral psychologists, the question of interest is not whether people are moral but whether people are motivated to see themselves as moral and, if so, how this motivation shapes behavior (e.g., Gino & Desai, 2012; Jordan & Monin, 2008; Monin & Miller, 2001).

The emerging consensus is that people strongly desire to affirm their moral identity—to maintain their moral self-regard by perceiving their own moral standing in a positive light (Aquino & Reed, 2002; Mazar, Amir, & Ariely, 2008; Monin & Jordan, 2009). And yet, acting morally often conflicts with people's other desires, such as maximizing their own economic self-interest. As a result, people typically do not seek to be as moral as they could be but simply "moral enough" (Miller & Monin, 2016; Nisan, 1990). Fortunately, people can affirm their moral identity without needing to be saints or even to take the most moral course of action in every situation (Lin, Zlatev, & Miller, 2017; Mazar et al., 2008; Merritt, Effron, & Monin, 2010; Ploner & Regner, 2013; Shalvi, Dana, Handgraaf, & De Dreu, 2011; Shalvi, Gino, Barkan, & Ayal, 2015). Instead, they are able to justify failing to take the costliest moral actions (e.g., Alleyne, Fernandes, & Pritchard, 2014; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Tsang, 2002), and even find ways to license their own clearly immoral behaviors (e.g., Monin & Jordan, 2009; Nisan, 1990; Sachdeva, Iliev, & Medin, 2009).

This portrait almost makes it appear as if the floor is the limit: As if people can feel they are moral actors no matter their behavior. But, surely, there is some minimum threshold: Some magnitude of good that a person's actions must yield if she is to affirm her moral self-regard. What is that threshold? More simply put, just how moral is moral enough? To begin to answer this question, we propose an account that we term the moral threshold model.

This article was published Online First November 11, 2019.

Julian J. Zlatev, Harvard Business School, Harvard University; Daniella M. Kupor, Questrom School of Business, Boston University; Kristin Laurin, Department of Psychology, University of British Columbia; Dale T. Miller, Graduate School of Business, Stanford University.

We thank Jen Dannals, Frank Flynn, Stephanie Lin, Nir Halevy, Christian Wheeler, and Robb Willer, as well as members of the DPER Lab at Stanford University, for their valuable feedback.

Correspondence concerning this article should be addressed to Julian J. Zlatev, Harvard Business School, Harvard University, Baker Library 443, Boston, MA 02163. E-mail: jzlatev@hbs.edu

This model accommodates existing findings on moral self-regard and generates novel predictions about how people make prosocial decisions.

The Moral Threshold Model

The moral threshold model identifies important parameters guiding people's morally-relevant decisions. Specifically, the moral threshold model offers two key hypotheses: the worst outcome avoidance hypothesis and the diminished upside sensitivity hypothesis.

The Moral Threshold Model in Two Key Hypotheses

The moral threshold model illuminates necessary and, in many cases, sufficient conditions that prosocial decisions must satisfy to affirm a person's moral self-regard. As in prior research, we define prosocial decisions as decisions that aim to help others in need (e.g., Chakroff & Young, 2015; Paciello, Fida, Cerniglia, Tramontano, & Cole, 2013; Suls, Witenberg, & Gutkin, 1981; cf. Choshen-Hillel & Yaniv, 2011; Shariff et al., 2016).

Worst outcome avoidance hypothesis. The first tenet of the moral threshold model is that, for a prosocial decision to affirm a person's moral self-regard, the decision's outcome must outshine at least one of the other outcomes that could result from the possible decision. This means that attempting but failing to do greater good is unlikely to affirm one's moral self-regard when a different available option would have guaranteed some good.[1] If the amount of good resulting from a choice is less than the amount of good guaranteed by other immediately available options, we propose that this option will do less to affirm the decision maker's moral self-regard than the other available options. If so, people may systematically avoid morally-relevant decisions that could produce the worst possible outcome in the available choice set, even when those decisions are objectively superior in other ways (e.g., having a higher expected value of benefit). This is the first key hypothesis of the moral threshold model: the *worst outcome avoidance* hypothesis.

This hypothesis suggests that, when it comes to moral self-regard, an old proverb is especially apt: a [morally enhancing] bird in the hand is worth more than two in the bush. For example, imagine Pam is deciding between either trying to help a group of homeless people find permanent housing (a risky undertaking which may succeed and provide a large long-term benefit, or fail and provide no benefit), or providing them with one guaranteed night of temporary shelter (guaranteed to provide a smaller, short-term benefit).[2] The moral threshold model posits that trying but failing to find permanent housing for these individuals—valiantly attempting to produce the best of all possible outcomes—is unlikely to provide a legitimate affirmation of Pam's moral self-regard, given that she has an alternative option. If Pam tries but fails to provide permanent housing, the outcome of her attempt to help—providing no benefit—is the worst outcome in her choice set. As a result, this outcome would not suffice to affirm her moral self-regard. Instead, the moral threshold model predicts that she is likely to make a decision which promises an outcome that compares favorably with at least one available outcome: In this case that is the decision to provide one guaranteed night of temporary shelter. Its benefit is smaller, but nonetheless compares favorably to the alternative where she tries, but fails, to provide the larger benefit.

Diminished upside sensitivity hypothesis. The moral threshold model's first hypothesis thus specifies the outcomes that are more versus less likely to affirm moral self-regard. Its second hypothesis specifies the exchange rate by which morally-relevant outcomes translate into moral self-regard. Specifically, it posits that this translation is nonlinear: As long as one has chosen an outcome that compares favorably to the worst available outcome—that is, as long as one has reached the minimum threshold specified in the worst outcome avoidance hypothesis—one gains little further moral self-regard from *how* favorable that comparison is. This is the second key hypothesis of the moral threshold model: the *diminished upside sensitivity* hypothesis.

This hypothesis cautions an addendum to the aforementioned proverb: Although a bird in the hand is worth more than two in the bush, once the first bird is in the hand there is little additional value in acquiring a second. This prediction goes beyond the well-known economic law of diminishing marginal utility (Gossen, 1983), which specifies the gradually decreasing marginal value of adding further gains on top of an initial gain. Specifically, the moral threshold model posits that the rate at which these marginal returns diminish is greater when making prosocial decisions than when making decisions for the self. In other words, Pam will place *even less* value on the successive nights of shelter she secured for the homeless than she would on successive nights of shelter she secured for herself. As long as the outcome of her choice compares favorably to one outcome from her choice set, Pam's claim of being a moral person is legitimated, and producing more benefit will add little to her experienced moral self-regard.

Existing Evidence in Support of the Moral Threshold Model

Some existing theory and research supports each of the hypotheses of the moral threshold model, and we detail this extant literature below.

Existing support for the worst outcome avoidance hypothesis. The worst outcome avoidance hypothesis posits that people have a liberal threshold for considering themselves moral: they must merely produce a magnitude of good greater than at least one available alternative. This liberal threshold is consistent with research revealing the scant basis that seems to suffice for people to feel morally affirmed (Bradley-Geist, King, Skorinko, Hebl, & McKenna, 2010; Effron, Monin, & Miller, 2013; Monin & Miller, 2001). Nevertheless, we maintain that a threshold does exist: People do require *some* evidence, however scant, to affirm their moral self-regard.

[1] The primary goal of the current research is to document people's *beliefs* about what kinds of decisions are likely to affirm their moral self-regard. People's beliefs about how they will feel guide their decisions (Zeelenberg, 1999), particularly when their self-image is involved (Josephs, Larrick, Steele, & Nisbett, 1992; Larrick, 1993), of which moral self-regard is a central component (Aquino & Reed, 2002). Of course, these beliefs may be inaccurate (e.g., Wilson & Gilbert, 2003, 2013); we return to this possibility in the General Discussion.

[2] Whereas the two options in this example might also vary in the amount of effort they require, this difference is not necessary for our predictions, and we hold effort constant in all of our studies.

This proposal aligns with work that shows the reality-constrained nature of motivated cognition (Dunning, Meyerowitz, & Holzberg, 1989; Kruglanski et al., 2012; Kunda & Sanitioso, 1989). People are flexible in their ability to feel good about their good behavior and to justify their bad behavior, but this flexibility is not unlimited. For example, employers make prejudiced evaluations of minority job candidates when the candidate's qualifications are ambiguous but not when the minority candidate is clearly the superior choice. In other words, an unambiguous external reality (a candidate's obviously stellar qualifications) makes it difficult for people to enact their prejudices (Merritt et al., 2012). In a similar vein, although prisoners demonstrate the better-than-average effect by believing they are kinder and more dependable than the average citizen, they do not claim to be more *law-abiding* than the average citizen. Here as well, an unambiguous external reality (prison life's constant reminder that one violated the law) makes it difficult for people to claim undeserved positive attributes (Sedikides, Meek, Alicke, & Taylor, 2014). Building on this work, we argue that the external environment in which a morally tinged decision is made can similarly constrain people's ability to conclude that they made a moral choice. If an individual's choice may produce the unambiguously worst prosocial outcome, it might be difficult for that person to lean on that choice to affirm her moral self-regard.

At first glance, these predictions appear to have two surprising implications. The first, that people may satisfice (rather than maximize) the amount of prosocial good they contribute, actually has strong empirical and theoretical support in the existing literature (e.g., Berman, Barasch, Levine, & Small, 2018; Mazar et al., 2008). The second, that people may anticipate that it would be difficult for them to claim moral self-regard on the basis of their intentions alone, may also at first glance appear to clash with theoretical perspectives suggesting that intentions feature prominently in moral judgments (Alicke, 2000; Malle, Guglielmo, & Monroe, 2014); however, the context we investigate differs substantially from those used in this prior research. After detailing our findings, we discuss how the current findings can be reconciled with these apparent contradictions.

Existing support for the diminished upside sensitivity hypothesis. The diminished upside sensitivity hypothesis posits that people are less sensitive to the magnitude of the possible upside payoff in a prosocial—compared with a purely self-interested—decision. This hypothesis adds an important codicil to the general claim that people can be insensitive to changes in magnitude (e.g., Hsee & Rottenstreich, 2004; Kahneman & Knetsch, 1992): Namely, that this insensitivity is greater when people make morally-relevant, versus morally-irrelevant, decisions. Prior literature finds that scope insensitivity exists in both types of decisions—for example, people will pay a similar amount of money for five CDs as they will for 10 CDs (Hsee & Rottenstreich, 2004), and will also pay a similar amount of money to clean up one polluted river as they will to clean up 10 polluted rivers (Mitchell & Carson, 1989). However, the moral threshold model predicts that people are *more* insensitive to potential upsides when making morally-relevant decisions (e.g., cleaning up rivers), than when making morally irrelevant decisions (e.g., buying CDs).

Also important, when prior work has examined potential scope insensitivity in prosocial decisions, it has operationalized these decisions as costing time or money—that is, as detracting from self-interested material goals (e.g., Imas, 2014). Doing so means that this prior work cannot identify what drove people's decisions: a burning desire to preserve finances or time after the minimum initial contribution, or a satiated desire to contribute moral good beyond that minimum contribution. People may have been unwilling to donate additional money for river clean-up not because it would have failed to meaningfully increase their moral self-regard, as we predict here, but because they preferred to spend this additional money elsewhere. We therefore test the moral threshold model (a) independent of any opportunity cost of spending one's resources elsewhere and (b) by contrasting exclusively prosocial decisions to decisions that exclusively benefit the self.

Summary and Overview of Research

To summarize, the moral threshold model predicts that people who make a morally-relevant decision will choose an option that guarantees a benefit that compares favorably with the worst possible outcome in the available choice set over one whose payoff has a higher expected value but carries the risk of producing the worst possible outcome. This is because any payoff that exceeds the worst possible outcome is sufficient to affirm the actor's moral self-regard, whereas a larger hoped-for payoff that fails to materialize is not. The model further predicts that, once this threshold is crossed, any additional amount of prosocial good that an action can produce has relatively little impact on people's decisions.

Six studies examine these predictions in the domain of risk, which allows for nuanced tests of our hypotheses while ruling out alternative explanations. Studies 1A and 1B find that people more strongly prefer decisions that guarantee outcomes which exceed the worst possible available outcome when making prosocial (vs. egoistic) decisions. Studies 2A and 2B rule out three alternative accounts: The predicted preference is not driven by (a) an asymmetry in the perceived number of people affected by prosocial versus self-interested decisions, (b) reputation maintenance concerns, or (c) a desire to specifically avoid contributing nothing to a prosocial cause. Study 3 provides evidence that the current phenomenon seems to emerge at least in part because prosocial decisions implicate moral self-regard. Finally, Study 4 finds that prosocial decisions are relatively insensitive to upside: When a benefit exceeds the worst possible available outcome, the magnitude by which this threshold is surpassed sways decisions less when the beneficiary is a prosocial cause rather than the self.

The risk context in which we test our predictions affords several advantages. First, as noted above, it equates the cost of all prosocial actions, thereby decoupling the value of the moral self-regard people acquire from producing a prosocial benefit from the personal cost of producing that benefit. That is, in each study we examine people's decisions when they face a choice between providing a certain, but small, donation to a prosocial cause versus an uncertain, but larger, donation to the same prosocial cause. The material cost of doing good is identical in both cases, but choosing the larger but riskier donation carries the psychic cost of potentially resulting in the smallest possible donation. Second, we compare prosocial risk decisions with purely egoistic ones that are otherwise identical. This allows us to determine whether the pattern of prosocial risk aversion predicted by the moral threshold model differs from the well-known risk aversion and certainty effects for decisions exclusively affecting the self (Kahneman &

Tversky, 1979; Rabin, 2000; Tversky & Kahneman, 1986). It also enables us to compare prosocial and egoistic choices that have the exact same structure in terms of their probabilities and expected values, ruling out the possibility that the specific values in our risk tradeoffs (which we vary across experiments) led people to avoid the risky choice.

Studies 1A, 1B, 2B, 3, and 4 were preregistered. In all studies, we followed the recommendations provided by Simmons, Nelson, and Simonsohn (2018) and collected at least 100 participants per condition. We limited our measures to only those that directly tested the two focal hypotheses. We report all relevant measures (i.e., all measures other than filler items) and no participants were excluded. All data, code, materials, and preregistrations are available at osf.io/38zy2/. All studies received approval from the relevant IRB.

Studies 1A and 1B

Studies 1A and 1B offer an initial test of the worst outcome avoidance hypothesis: that when people make decisions pertaining to the welfare of others in need (vs. purely egoistic decisions), they are more likely to choose an option that ensures an outcome that compares favorably to at least one other possible outcome. Study 1A examines decisions about hypothetical monetary investments; Study 1B puts participants' real money at stake. Both Study 1A and Study 1B were preregistered.

Study 1A: Method

Participants. Five hundred three participants (43.8% female; $M_{age} = 34.42$, $SD_{age} = 11.03$; one participant did not complete the demographic questions) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to either a *Prosocial* condition or an *Egoistic* condition. In the *Prosocial* condition, participants imagined that they worked at a charity, and that this charity had recently received a $200 gift. Participants learned that they were in charge of deciding what to do with this money. Specifically, they learned that they were deciding between two options: If they selected Option A (hereafter referred to as the sure option), they would put the money into the charity's savings account, in which case they would be able to ensure that the charity kept the $200; if they selected Option B (hereafter referred to as the risky option), they would put the money into an investment in which the charity had a 75% chance of receiving $275 and a 25% chance of receiving $0 (i.e., losing the $200). Thus, the risky option had a higher expected value ($206.25) than the sure option ($200), but the risky option also carried the possibility of yielding the worst outcome (i.e., the smallest donation) in the choice set. By contrast, the sure option guaranteed a donation that compared favorably to this worst alternative.

Participants in the *Egoistic* condition made the same risk decision, but instead imagined that they had themselves received a gift of $200, and that they were deciding what to do with this money. Participants then viewed the same two options that participants in the *Prosocial* condition viewed.

In both conditions, participants entered their choice by selecting a radio button labeled either "Option A" or "Option B." As a supplemental measure, participants also responded to a continuous measure of which option they preferred on a 7-point scale, which ranged from 1 (*definitely prefer option A*) to 7 (*definitely prefer option B*).

Study 1A: Results and Discussion

As predicted, a binomial logistic regression revealed that participants in the *Prosocial* condition chose the risky option—the option that carried the risk of yielding the worst possible outcome—less often (10.4%) than participants in the *Egoistic* condition (24.0%; $\chi^2[df = 1, N = 503] = 15.26$, $p < .001$, $\phi = .18$). A t test examining the continuous measure corroborated this finding: Participants in the *Prosocial* condition ($M = 2.26$, $SE = .11$) were more averse to the risky option than participants in the *Egoistic* condition ($M = 2.88$, $SE = .13$; $t[501] = 3.73$, $p < .001$, Cohen's $d = .33$). In other words, people were more hesitant to make a decision that could result in the worst possible outcome in the choice set when that decision was morally-relevant rather than purely self-interested, even when that meant foregoing an objectively better option (i.e., the option with the larger expected value). This result is consistent with the worst outcome avoidance hypothesis (i.e., the first hypothesis of the moral threshold model).

Study 1B: Method

Participants. Five hundred participants (42.4% female; $M_{age} = 34.38$, $SD_{age} = 11.69$) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. All participants first completed a filler task asking them to evaluate six images, and then faced a decision about real money. In the *Prosocial* condition, participants made a real donation decision for a charity of their choice. Specifically, participants were presented with three charities (United Way, Habitat for Humanity, and the American Red Cross) that they could choose to donate to. After choosing a charity, participants learned that they could select one of two donation options: Option A (the sure option) guaranteed that their chosen charity would receive 25 cents; in contrast, Option B (the risky option) offered a 75% chance that their chosen charity would receive 45 cents, and a 25% chance that the charity would receive no money (i.e., an outcome smaller than the amount guaranteed by the sure option). Participants read that if they selected the risky option, a random number generator would determine its outcome. Thus, the risky option had a higher expected value ($0.34) than the sure option ($0.25). The instructions emphasized that this was a real donation decision, and that participants' decision would determine how much money would actually be donated to each participant's selected charity.

Participants in the *Egoistic* condition made the same real decision, but learned that they would receive the money resulting from their decision as a bonus payment. They then made the same risk decision as participants in the *Prosocial* condition. As in the *Prosocial* condition, the instructions emphasized that this was a real decision, and that their decision would determine how much money they would receive as a bonus.

In both conditions, participants entered their choice by selecting a radio button labeled either "Option A" or "Option B." After making their decision, participants who chose the risky option were shown the results of the random number generator that determined how much they or the charity would receive.

Study 1B: Results and Discussion

As predicted, participants in the *Prosocial* condition were less likely to choose the risky option (33.2%) than were participants in the *Egoistic* condition (54.4%; $\chi^2[df = 1, N = 500] = 21.97$, $p < .001$, $\phi = .21$). In sum, even in a choice setting with real stakes, people were more averse to choosing options that carry the risk of yielding the worst possible outcome in the choice set when making prosocial decisions than when making self-interested decisions.[3]

Studies 2A and 2B

The previous studies support the worst outcome avoidance hypothesis. First, these studies demonstrate that people prefer prosocial decision options that guarantee an amount that is larger than the worst possible available outcome (even if that guaranteed outcome is small) over options that have a larger expected value but that could result in the worst possible outcome. Moreover, Studies 1A and 1B found that this preference is larger in prosocial decision contexts than egoistic decision contexts. Together, these findings are consistent with the moral threshold hypothesis's prediction that people can only affirm their moral self-regard if their prosocial decisions produce outcomes that are better than the worst outcome in the available choice set. Studies 2A and 2B provide additional tests of this hypothesis, while also examining potential alternative explanations.

Specifically, Study 2A examines the possibility that the worst prosocial outcome avoidance documented in the prior studies is simply due to the fact that the prosocial (vs. egoistic) outcomes in the prior studies affected more individuals. Study 2B examines the possibility that worst prosocial outcome avoidance results from a motive to maintain one's public reputation rather than one's private moral self-regard. Study 2B also examines the possibility that the prior studies may not have documented worst prosocial outcome avoidance, but rather a more specific motive to avoid causing precisely zero prosocial good.

Study 2A

The prosocial condition in each of the previous studies featured numerous beneficiaries (i.e., all of the individuals served by a given a charitable organization), while the egoistic beneficiary in each of the previous studies was a single individual (i.e., the participant). As a result, it is possible that we found an asymmetry in these previous studies because people think differently about multiple versus single beneficiaries, and not, as we hypothesized, because people think differently about beneficiaries in need than about their own self-interest.

To address this alternative explanation, Study 2A examines whether the previously documented worst prosocial outcome avoidance persists when people make a prosocial decision for a single beneficiary in need: an individual homeless person. If people's avoidance of the worst possible outcome depends on the number of potential beneficiaries, it should not emerge in Study 2A because the decisions in all conditions impact a single beneficiary. In contrast, the moral threshold model predicts that participants making a risk decision that benefits even a single individual in need (vs. themselves) will continue to be more averse to decisions that carry the risk of producing the worst possible outcome.

Study 2A also seeks to reconcile our findings with previous research which finds that people choose decision options for others that are either riskier than (Andersson, Holm, Tyran, & Wengström, 2016; Polman, 2012; Wray & Stone, 2005), or just as risky as (Stone, Yates, & Caruthers, 2002), those that they choose for themselves. We contend that the aversion to risky decisions we observed in our earlier studies is specific to risk decisions that help not just any others, but others in need (i.e., that have prosocial implications). To test this prediction, in Study 2A we include a third condition in which participants make a decision on behalf of an individual who is neither the self nor in need.

Study 2A: Method

Participants. Four hundred forty-nine participants (40.9% female; $M_{age} = 33.10$, $SD_{age} = 10.94$; four participants did not complete the demographic questions) from Amazon's Mechanical Turk took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to a *Prosocial* (*other in need*) condition, a *Non-Prosocial (other not in need*) condition, or an *Egoistic* condition. In all conditions, participants imagined that they were making a choice between two options: Option A and Option B. Option A (the sure option) offered a sure $5. In contrast, Option B (the risky option) offered a 75% chance of $9, and a 25% chance of $0 (i.e., an amount less than the $5 guaranteed by the sure option). Thus, the risky option had a greater expected value ($6.75) than the sure option ($5.00). The three conditions differed only in the recipient that participants were told would receive the money accrued from their decision—participants in the *Prosocial (other in need*) condition were told that a homeless individual would receive the money, participants in the *Non-Prosocial (other not in need*) condition were told that another survey participant would receive the money, and participants in the *Egoistic* condition were told that they themselves would receive the money. As in the previous studies, participants entered their choice by selecting a radio button labeled either "Option A" or "Option B."

Study 2A: Results and Discussion

A chi-square analysis revealed a significant effect of condition on risk choices, $\chi^2(df = 2, N = 449) = 11.54$, $p = .003$, $\phi = .16$. Consistent with our prediction, further analyses found that participants who made a decision for a single individual in need (i.e., a prosocial decision) chose the risky option significantly less frequently (11.3%) than did participants who made a decision for

[3] Although Study 1B used real financial stakes, this is only meaningful if participants believed that the stakes were in fact real. As a robustness check, we therefore asked participants to indicate, immediately after their decision, if they were suspicious about whether anything they had read so far in the study was not true. The majority of participants (68.2%) selected "No, I was not suspicious" (rather than "Yes, I was suspicious"). Results held when we excluded suspicious participants: Participants in the *Prosocial* condition still chose the risky option less often (33.1%) than participants in the *Egoistic* condition (56.0%; $\chi^2[df = 1, N = 341] = 16.94$, $p < .001$, $\phi = .23$). Thus, Study 1B provided evidence consistent with the worst outcome avoidance hypothesis even when participants *knew* real money was at stake.

themselves (21.8%; $\chi^2[df = 1, N = 298] = 5.25$, $p = .022$, $\phi = .14$) or for another single individual not in need (26.5%; $\chi^2[df = 1, N = 302] = 10.47$, $p = .001$, $\phi = .19$). The latter two conditions did not differ significantly from one another, ($\chi^2(df = 1, N = 298) = 0.67$, $p = .41$, $\phi = .06$). This null effect aligns with work finding no difference between monetary risk decisions for others versus the self (e.g., Stone et al., 2002); nevertheless, its (nonsignificant) direction mirrors that sometimes observed in other prior literature (Andersson et al., 2016; Polman, 2012; Wray & Stone, 2005).

This study yields further evidence consistent with the moral threshold hypothesis that people have a stronger desire to guarantee an amount that is larger than the worst possible outcome when the decision has prosocial implications compared to when it does not (e.g., when the decision is for one's own profit or for the profit of a non-needy stranger).

Study 2B

Study 2B examines two additional alternative accounts for the results documented thus far. The first is that people may avoid the worst prosocial outcome to maintain their public reputation—that is, to promote others' moral regard for them, rather than to affirm their own private moral self-regard. This possibility seems unlikely, given that participants completed the previous studies online and anonymously, and nothing in the instructions suggested that their decisions or their outcomes would be broadcast to others. Nevertheless, in Study 2B, we randomly assigned participants to a public condition (in which we made it clear that their decision and its outcome would be publicly known) or a private condition (in which we made it clear that their decision and its outcome would remain anonymous). Because the moral threshold model concerns people's desire to maintain their *private* moral self-regard, we predicted that—even when participants learned that no one would ever know their decision or its result—they would continue to avoid the risky option when making decisions for others in need. Of course, we also expected that participants would similarly avoid the prosocial risky option when they made their decision in public, because participants' private desire to affirm their moral self-regard should guide their prosocial decisions not only in private but also in public.

The second alternative account that Study 2B investigates is that participants may not prefer to avoid *any* worst possible outcome, but rather the specific outcome of causing zero prosocial good. For that reason, and in contrast to Studies 1A–2A (in which the worst outcome was always zero), in Study 2B we set the worst possible outcome to be nonzero (see also Footnote 4). Study 2B was preregistered.

Study 2B: Method

Participants. Six hundred three participants (48.2% female; $M_{age} = 33.09$, $SD_{age} = 11.38$; three participants did not complete demographic questions) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to one condition in a 2 (Decision Beneficiary: Egoistic vs. Prosocial) × 2 (Decision Context: Private vs. Public) design. In all conditions, participants imagined that they were billionaires, and that they were making a choice between two options: Option A and Option B. Option A (the sure option) offered a sure $1,000,000. In contrast, Option B (the risky option) offered a 75% chance of $1,800,000, and a 25% chance of $40.[4] Thus, the risky option had a greater expected value ($1,350,010) than the sure option ($1,000,000).

As in the prior studies, we manipulated the beneficiary of this decision: Participants in the *Prosocial* condition learned that a charity would receive the money, whereas participants in the *Egoistic* condition learned that they themselves would receive the money. We also manipulated the private versus public nature of their decision via a manipulation similar to those used in prior research (e.g., Kraus & Callaghan, 2016; White, Simpson, & Argo, 2014): Participants in the *Private* condition read that no one else would ever learn about their decision or their decision's outcome, whereas participants in the *Public* condition read that many others would learn about their decision as well as their decision's outcome.

As in the previous studies, participants entered their choice by selecting a radio button labeled either "Option A" or "Option B."

Study 2B: Results and Discussion

We conducted a 2 (Decision Beneficiary: Egoistic vs. Prosocial) × 2 (Decision Context: Private vs. Public) binary logistic regression on participants' decisions. As predicted, the regression revealed a main effect of decision context: Participants in the *Prosocial* conditions were less likely to choose the risky option (13.3%) than were participants in the *Egoistic* conditions (39.2%; $b = 1.43$, $z = 6.93$, $p < .001$). Also as predicted, there was no significant interaction between decision beneficiary and decision context ($b = .10$, $z = .24$, $p = .81$): Participants in the *Prosocial* condition chose the risky option less frequently than did participants in the *Egoistic* condition both when their decision was private (*Prosocial* condition = 15.2%, *Egoistic* condition = 42.0%, $b = 1.49$, $z = 4.84$, $p < .001$) and when it was public (*Prosocial* condition = 11.4%, *Egoistic* condition = 36.4%, $b = 1.39$, $z = 4.97$, $p < .001$). The regression also revealed no difference in choice of the risky option in the *Private* condition (28.6%) compared to the *Public* condition (24.0%; $b = .24$, $z = 1.27$, $p = .20$; see Figure 1).

These findings support our contention that it is people's motivation to maintain their private moral self-regard—not motivation to maintain their public reputation—that fuels their aversion to making a decision that could result in the worst possible outcome for others in need. Moreover, because this worst possible outcome provided a net positive prosocial good, these findings also rule out

[4] We acknowledge that, although technically not zero, the worst possible outcome in Study 2B's choice set was very small relative to the other possible outcomes, and participants could thus have encoded it as essentially zero. Therefore, in Study 3, we use a nonzero outcome that is less paltry in comparison to the other possible outcomes; we also ran a supplemental study specifically aimed to address this. In the supplemental study, participants ($N = 300$) decided between a guaranteed $200 versus a 75% chance of $275 and a 25% chance of $60. Participants were randomly assigned to either a *Prosocial* or *Egoistic* condition using the methodology described in Study 1A. Replicating previous studies, participants chose the risky option significantly more frequently in the *Egoistic* condition (31.6%) than in the *Prosocial* condition (13.8%; $\chi^2[df = 1, N = 300] = 12.45$, $p < .001$, $\phi = .21$), even when the risky option guaranteed some benefit.

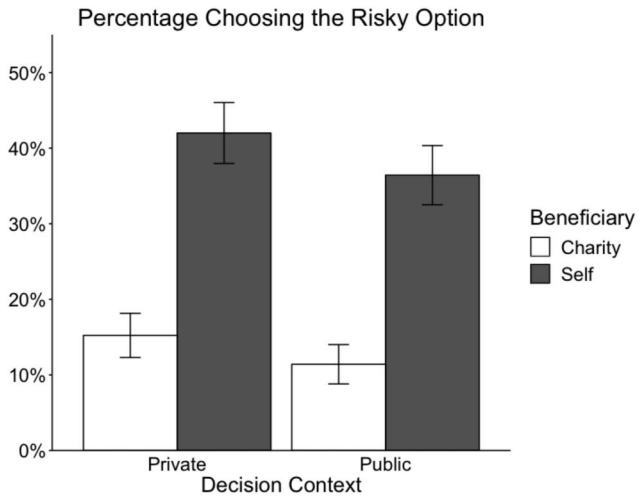

Figure 1. Percentage of participants in Study 2b choosing the risky option as a function of beneficiary type and decision context.

the possibility that people are specifically motivated to avoid contributing exactly zero to a prosocial cause (see also footnote 4 for a replication of this finding using a comparatively more substantial nonzero worst possible outcome). Study 3 tests both of these notions further.

Study 3

Study 3 uses mediation to test whether people avoid decisions that yield the worst possible prosocial outcome in the available choice out of a concern for their moral self-regard. Specifically, after participants made a decision either for a prosocial cause or for themselves, they reported the concern they would have for their moral self-regard if they had chosen the risky option and it resulted in the smallest available outcome. We predicted that this concern would be greater, and would account for participants' greater worst prosocial outcome avoidance, when they made a prosocial decision rather than a purely self-interested decision. Moreover, in Study 3 the worst available outcome again guaranteed a positive prosocial good, and this time represented a much larger fraction of the alternative outcomes than was the case in Study 2B. Study 3 was preregistered.

Method

Participants. Six hundred one participants (47.9% female; M_{age} = 33.22, SD_{age} = 11.99; eight participants did not complete the demographic questions) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. All participants imagined that they were making a choice between two options: Option A and Option B. Option A (the sure option) offered a sure $20. In contrast, Option B (the risky option) offered a 75% chance of $40, and a 25% chance of $4. Thus, the risky option had a greater expected value ($31) than the sure option ($25). The two conditions differed only in the recipient that participants were told would receive the money accrued from their decision—participants in the *Prosocial* condition were told that a charity would receive the money, whereas participants in the *Egoistic* condition were told that they themselves would receive the money. As in the previous studies, participants entered their choice by selecting a radio button labeled either "Option A" or "Option B."

After participants entered their risk decision, they advanced to the next page and completed a two-item index assessing the degree to which their decision reflected concern for their moral self-regard. They imagined that they had selected the Option B (the risky option) and that as a result either they or their chosen charity had received $4 (the worst available outcome). Participants then answered the following two questions on a 7-point scale (1 = *not at all*; 7 = *very much*): "To what extent would you feel like a bad person?" and "To what extent would you feel less moral than you would like to feel?" We combined these items (r = .77) into an index assessing the extent to which participants perceived the decision implicated their moral self-regard.

Results and Discussion

Replicating the previous studies, participants in the *Prosocial* condition were less likely to choose the option that yielded a higher expected value but carried the risk of yielding the worst possible benefit (33.6%) than were participants in the *Egoistic* condition (47.3%; $\chi^2[df = 1, N = 601] = 11.28$, $p < .001$, $\phi = .14$).

Also as predicted, the decision in the *Prosocial* condition prompted greater concern for moral self-regard (M = 3.54, SD = 2.46) than did the decision in the *Egoistic* condition (M = 2.15, SD = 1.65; $t(599) = 8.14$, $p < .001$, Cohen's d = .66). This greater concern accounted for the difference in participants' risk decisions: A mediation analysis with bootstrapping (10,000 iterations) following the recommendation of Hayes (2013) in which we entered condition as the independent variable, participants' decisions as the dependent variable, and participants' concern for their moral self-regard as the mediator revealed a significant indirect effect (95% CI [.052, .114], see Figure 2).

In sum, consistent with the moral threshold model, people feel that a choice that yields the worst possible outcome threatens their moral self-regard, which in turn predicts their tendency to avoid choices that carry the risk of that outcome. Of course, cross-sectional mediation analyses such as this do not warrant definitive causal claims; nevertheless, the current results are consistent with the possibility that people's concerns for their moral self-regard

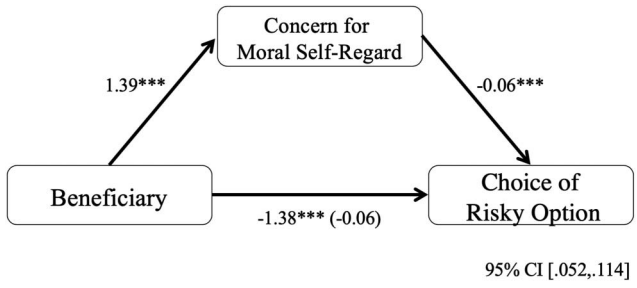

Figure 2. Mediation by concern for moral self-regard in Study 3. The path coefficients are unstandardized betas. The value in parentheses indicates the effect of beneficiary condition on the dependent variable after controlling for the mediator. *** $p < .001$.

drive the worst prosocial outcome avoidance we have documented here (see Shrout & Bolger, 2002).

Study 4

Study 4 tests the diminished upside sensitivity hypothesis, which proposes that when making prosocial (vs. purely egoistic) decisions, people are less sensitive to the magnitude of benefit they will potentially provide (so long as that magnitude exceeds the worst possible outcome in the choice set). Study 4 was preregistered.

Method

Participants. Four hundred three participants (48.0% female; $M_{age} = 31.01$, $SD_{age} = 11.21$) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. Participants imagined that they had received a gift of $200 and were deciding what to do with the money. Similar to the previous studies, participants were randomly assigned to either an *Egoistic* condition (in which the money was for them) or a *Prosocial* condition (in which the money was for a charity of their choice). Participants then made decisions in six separate scenarios; in each one, Option A (the sure option) guaranteed $200, and Option B (the risky option) offered a 75% chance of an amount larger than $200 but a 25% chance of $0. The six scenarios varied only in the larger amount of money that could result from the risky option: This amount varied in $10 increments between $250 and $300, and for each participant the amounts appeared in either increasing or decreasing order through the scenarios. Participants entered their choice by selecting a radio button labeled either "Option A" or "Option B" before moving on to the next scenario.

The worst outcome avoidance hypothesis predicts that participants would more greatly prefer the safe option when they chose for a prosocial cause than when they chose for themselves. More importantly, the diminished upside sensitivity hypothesis predicts that participants would be less tempted by the risky option as its larger potential outcome increased when they chose for a prosocial cause than when they chose for themselves.

Results and Discussion

A mixed-effects binary logistic regression, treating participant as a random effect, revealed a main effect such that participants were more likely to choose the risky option as its upside increased ($b = .98$, $z = 12.20$, $p < .001$). More importantly it revealed an interaction consistent with the diminished sensitivity hypothesis ($b = -.41$, $z = 3.14$, $p = .002$). As predicted, simple slopes analyses revealed that participants' decisions were more influenced by the risky option's upside amount in the *Egoistic* condition ($b = 1.17$, $z = 11.31$, $p < .001$) than in the *Prosocial* condition ($b = 0.77$, $z = 8.14$, $p < .001$; see Figure 3).

Although not the focus of Study 4, the worst outcome avoidance hypothesis also received support in this same analysis: Participants who were making their decisions for a prosocial cause were less likely to choose the risky option than participants who were making the decision for themselves ($b = -1.45$, $z = 4.02$, $p < .001$).

After completing our preregistered analyses, we conducted an exploratory analysis to examine whether the order in which we presented the scenarios affected our key results. In a model that included presentation order as a factor, there was no three-way interaction ($p = .94$), and the Condition × Upside Amount interaction predicted by the diminished upside sensitivity hypothesis emerged within each presentation order (decreasing order: $p = .02$; increasing order: $p = .02$). Unexpectedly, however, there was an interaction between presentation order and condition ($p < .001$): the worst outcome avoidance hypothesis received support when

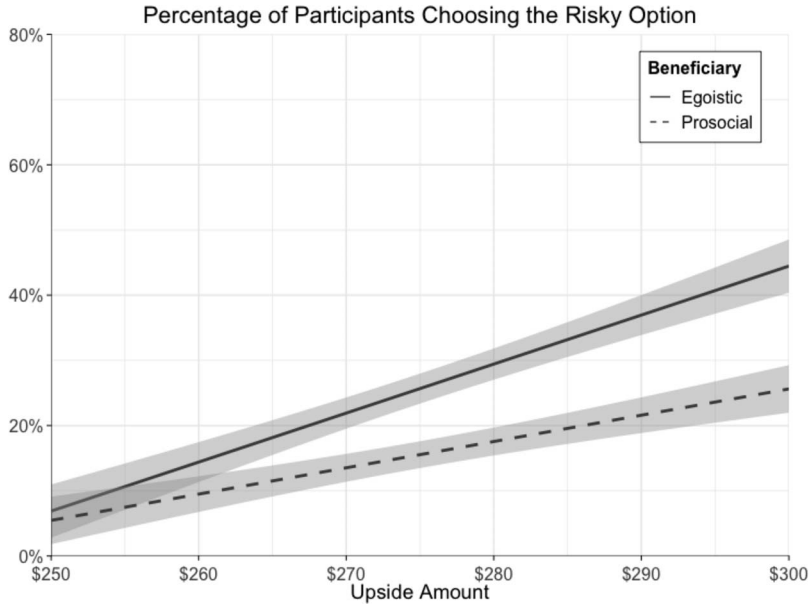

Figure 3. Percentage of participants in Study 4 choosing the risky option as a function of beneficiary type and the risky option's upside amount.

the upside amounts were presented in decreasing order ($p < .001$) but not when they were presented in increasing order ($p = .25$). However, because the worst outcome avoidance hypothesis replicated in our other studies, as well as in the preregistered analysis for this study, we do not discuss this difference further.

Study 4 supports the second postulate of the moral threshold model—when people's moral self-regard is at stake, their choices are less sensitive to *how much* good they might do once they have produced an amount of good that is greater than the worst possible benefit in the choice set. Of course, participants in the prosocial condition were not *completely* insensitive to the risky option's upside: They were still more likely to choose risky options with larger potential upsides than risky options with smaller potential upsides, but this effect was significantly smaller than it was in the egoistic condition. As with most psychological models, the moral threshold model does not claim to account for every possible factor that individuals might consider when formulating their decisions. Instead, it merely predicts, as we have found here, that when people's moral self-regard is at stake, their choices are less sensitive to the magnitude of good they produce once their contribution reaches the initial threshold specified by the worst outcome avoidance hypothesis.

General Discussion

People select actions that allow them to maintain their moral self-regard (Aquino & Reed, 2002; Mazar et al., 2008; Monin & Jordan, 2009). But how exactly do moral actions translate into moral self-regard? Our moral threshold model proposes that when a decision has moral implications, people are primarily concerned with producing an outcome that will compare favorably with at least one alternative outcome in the available choice set. This translates into two hypotheses: (a) People will avoid choices that carry the possibility of resulting in the worst available prosocial outcome, and (b) so long as people have avoided this worst outcome, they will be relatively insensitive to the possibility of a larger upside.

In six studies, participants chose between opportunities to benefit a prosocial cause or themselves that varied in their outcomes' probability and magnitude but imposed the same cost on people's time, effort, and finances. In contrast with past scope insensitivity work, which used nonrisky contexts and conflated prosocial good with personal cost, this design allowed us to isolate which actions most strongly satiate people's desire for moral self-regard, separate from their material self-interest. Supporting the first prediction of the moral threshold model, we found that people showed a stronger preference for smaller guaranteed positive outcomes over larger uncertain ones when the beneficiary was a prosocial cause than when it was themselves (Studies 1A and 1B, 2A and 2B, 3, and 4). Supporting the second prediction of the moral threshold model, Study 4 showed that people's decisions are less sensitive to the size of the maximum potential benefit when making decisions for a charity, compared to when making decisions exclusively for themselves. Table 1 summarizes the primary results from each study.

In short, these findings support our contention that people are motivated to provide a benefit that exceeds the worst possible available outcome in order to affirm their moral self-regard, but are relatively unmotivated to provide further contributions because those contributions have little additional impact on their moral self-regard.

Table 1
Summary of Main Effect of Beneficiary (i.e., Egoistic vs. Prosocial) Across Studies 1–4

	Percent choosing risky option by condition	
Study	Egoistic	Prosocial
1a	24.0%	10.4%
1b	54.4%	33.2%
2a	21.8%	11.3%
2b	39.2%	13.3%
3	47.3%	33.6%
4	25.7%	15.5%

Note. Study 2b presents results collapsed across *Public* and *Private* conditions. Study 4 presents results averaged across upside amount.

Reconciling Prior Work

The moral threshold model concerns how people make decisions in advance of their outcomes. By design, then, it relies on people's expectations of how they will feel about these decisions' various possible outcomes (Zeelenberg, 1999). Past work has outlined the myriad ways by which people justify their prior decisions in hindsight, after the consequences of their decisions are known (Kunda, 1990; Shalvi et al., 2015). As such, although our research suggests people anticipate that producing the worst available prosocial outcome would threaten their moral self-regard, it may be that people who actually *have* produced that outcome would find ways to defend against that threat. For example, they might attribute the outcome to self-irrelevant causes, or console themselves with the thought that at least they tried to do some good (Shalvi et al., 2015).

This last possibility highlights one possible resolution to the aforementioned apparent conflict between the moral threshold model and past work on the focal role of intentionality in guiding judgments of morality (Alicke, 2000; Malle et al., 2014). More specifically, on the one hand, our model and findings suggest that people do not expect that good intentions—at least in the form of an unrealized prosocial outcome of larger expected value—will affirm their moral self-regard. On the other hand, prior literature finds that people use intentions as a critical cue of whether or not an action is morally permissible (Cushman, 2008). Although at first glance the current findings may appear at odds with this prior work, it is possible that people do not *expect* good intentions alone to affirm their moral self-regard, but once intentions have failed to materialize, they become fodder for justification.

There are also additional ways in which our findings differ from this prior work. Whereas our studies concerned judgments about *one's own moral* behavior (i.e., self-perceptions of decisions to help), most of the research on intentions concerns judgments about the *immoral* behavior of *others* (i.e., third-party perceptions of decisions to harm; e.g., Cushman, Young, & Hauser, 2006; Greene et al., 2009; cf. Berthoz, Grèzes, Armony, Passingham, & Dolan, 2006; Young & Saxe, 2011). It is possible that the focus of one's judgment (i.e., self or other), the type of behavior (i.e., moral or immoral), or both may affect the role of intentions.

Finally, and perhaps most importantly, prior work has primarily focused on contexts in which the alternative to behaving prosocially was to instead behave self-interestedly. In other words, having good intentions in these contexts compared favorably with the available alternative of *not* having good intentions at all (e.g., Kruger & Gilovich, 2004). This prior work may therefore be consistent with our worst outcome avoidance hypothesis in that the worst option in this prior literature (i.e., having no prosocial intentions and instead choosing the self-interested option) was simply even worse (i.e., less moral) than the options examined in the current research.

How Much Moral Self-Regard Do People Want?

The second hypothesis of the moral threshold model indicates that people are not primarily concerned with producing the *most* possible prosocial good, but rather with merely surpassing a minimum threshold. Future research might disentangle two possible mechanisms for this hypothesis, and in so doing might shed light on the inner workings of moral self-regard itself. The first possibility is that once people's actions have exceeded the minimum threshold, moral self-regard itself is then insensitive to further increases above that threshold. According to this possibility, when people produce a benefit that exceeds the worst possible outcome by 10 units rather than by one unit, they gain relatively little additional moral self-regard: In both cases they see themselves as identically moral. A second possibility is that moral self-regard actually is sensitive to magnitude, but that greater feelings of moral self-regard beyond some specific amount provide little additional psychological value. According to this possibility, people do see themselves as more moral if they exceed the worst possible outcome by 10 units rather than by one unit, but they simply do not care about this difference; exceeding the threshold by one unit might make them maximally satisfied with their moral self. Our studies do not adjudicate between the two possible accounts for the scope insensitivity we observed.

Both of these possibilities evoke the traditional distinction between satisficing and maximizing (Simon, 1956); in other words, our findings suggest that people may tend to satisfice, rather than maximize, in the moral domain. At first glance, this may be surprising given that the moral domain is one that people care deeply about (Aquino & Reed, 2002). And yet, as previously mentioned, caring deeply about being moral does not ensure that people will always do the right thing (Mazar et al., 2008). Indeed, we are not the first to suggest people can achieve a sufficient level of moral self-regard without the need to do the most good possible in a given situation (e.g., Berman et al., 2018).

Boundaries of the Moral Threshold Model

We have neither claimed nor demonstrated that actions that provide a prosocial benefit will always be evaluated through a moral lens, and that decisions with prosocial implications will thus always conform to the moral threshold hypothesis. The moral lens we identify is not the only one through which people may view their morally-relevant actions. For example, contrast a regular civilian who is fundraising for a charity with a person whose job it is to fundraise for that same charity. We propose that the regular civilian will accrue little additional benefit if she solicits a donation that exceeds the smallest possible donation by $1 rather than by $10, because her reactions to these outcomes will be rooted in her success in maintaining her moral self-regard. In contrast, the professional fundraiser may not use the lens of moral self-regard, but rather the lens of material utility, to evaluate her work. As a result, she may feel substantially more satisfaction from the donations she attracts as they get larger and larger.

We have also neither claimed nor demonstrated that the moral threshold model applies as a human universal. A key assumption of our work is that people desire to feel that they are moral, and that this desire drives behavior in morally-relevant decision making. Whereas the present work employed samples that included participants with a broad range of ages and socioeconomic statuses within the United States, cross-cultural work has questioned whether positive self-regard is indeed a universal motivation beyond Western cultures (Heine, Lehman, Markus, & Kitayama, 1999). As such, it may be the case that the desire for moral regard is diminished in more interdependent cultures, which would mitigate the effects found here. Future work should examine whether cultural interdependence moderates the findings documented here, or whether different patterns of results emerge among different cultures.

In addition to examining the moral threshold hypothesis in a wider range of cultures, future research might also employ a wider range of stakes beyond the relatively small ones used in the present studies. It is important to note, however, that Study 1B documented the current phenomenon in the context of real payouts of amounts that generate decisions comparable to the decisions made in response to larger stakes (Amir, Rand, & Gal, 2012). Moreover, research investigating egoistic risk decisions reveals similar effects across a wide variety of stakes (and, if anything, reveals that higher stakes only increase the risk aversion that emerges with smaller stakes; e.g., Holt & Laury, 2002). Still, future research could benefit from examining how the stakes in a risky context affect prosocial risk decisions.

Finally, we encourage future research to explore whether additional processes beyond those documented in the current research may contribute to the phenomenon we have identified. Although we find evidence consistent with our contention that people's concern for their moral self-regard underlies the documented effect, psychological phenomena are typically multiply determined (Kirmani, 2015); thus, our theorizing does not preclude the possibility that other factors may also contribute to our results. For example, perhaps the current phenomenon also emerges in part because choosing for others in need heightens feelings of responsibility, in turn increasing people's desire to avoid personally producing an undesirable outcome. Future research could profit from providing direct insight into this possibility.

Conclusion

Prior research has illuminated a considerable amount about how people respond to opportunities that invoke their self-interest. But how do people respond to opportunities that invoke their moral self-regard? The current work supports a moral threshold hypothesis, which proposes that when faced with a set of options with prosocial consequences, people are primarily concerned with ensuring they have a moral footprint greater than that of the worst possible outcome in the available choice set, but are less concerned with maximizing it. Interestingly, the current findings reveal that

people's desire to maintain their moral self-regard may ironically lead them to do less good. These results shed new light on how people make decisions that implicate their moral identity.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126,* 556.

Alleyne, E., Fernandes, I., & Pritchard, E. (2014). Denying humanness to victims: How gang members justify violent behavior. *Group Processes & Intergroup Relations, 17,* 750–762. http://dx.doi.org/10.1177/1368430214536064

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the internet: The effect of $1 stakes. *PLoS ONE, 7,* e31461. http://dx.doi.org/10.1371/journal.pone.0031461

Andersson, O., Holm, H. J., Tyran, J. R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preferences or noise? *Journal of the European Economic Association, 14,* 1129–1154. http://dx.doi.org/10.1111/jeea.12179

Aquino, K., & Reed, A., II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83,* 1423–1440. http://dx.doi.org/10.1037/0022-3514.83.6.1423

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71,* 364–374. http://dx.doi.org/10.1037/0022-3514.71.2.364

Batson, C. D. (2016). Moral motivation. In J.-W. van Prooijen & P. A. M. van Lange (Eds.), *Cheating, corruption, and concealment: The roots of dishonesty* (pp. 15–32). Cambridge, UK: Cambridge University Press.

Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to effective altruism: The role of subjective preferences in charitable giving. *Psychological Science, 29,* 834–844. http://dx.doi.org/10.1177/0956797617747648

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *Neuroimage, 31,* 945–950.

Bloom, P. (2012). Religion, morality, evolution. *Annual Review of Psychology, 63,* 179–199. http://dx.doi.org/10.1146/annurev-psych-120710-100334

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., & McKenna, C. (2010). Moral credentialing by association: The importance of choice and relationship closeness. *Personality and Social Psychology Bulletin, 36,* 1564–1575. http://dx.doi.org/10.1177/0146167210385920

Chakroff, A., & Young, L. (2015). The prosocial brain: Perceiving others in need, and acting on it. In L. M. Padilla-Walker & G. Carlo (Eds.), *Prosocial development: A multidimensional approach* (pp. 90–111). New York, NY: Oxford University Press.

Choshen-Hillel, S., & Yaniv, I. (2011). Agency and the construction of social preference: Between inequality aversion and prosocial behavior. *Journal of Personality and Social Psychology, 101,* 1253–1261. http://dx.doi.org/10.1037/a0024557

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108,* 353–380.

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.

De Waal, F. (2009). *Primates and philosophers: How morality evolved*. Princeton, NJ: Princeton University Press.

Dunning, D., Meyerowitz, J., & Holzberg, A. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090. http://dx.doi.org/10.1037/0022-3514.57.6.1082

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology, 49,* 573–578. http://dx.doi.org/10.1016/j.jesp.2012.08.012

Gino, F., & Desai, S. D. (2012). Memory lane and morality: How childhood memories promote prosocial behavior. *Journal of Personality and Social Psychology, 102,* 743–758. http://dx.doi.org/10.1037/a0026565

Gossen, H. H. (1983). *The laws of human relations and the rules of human action derived therefrom*. Cambridge, MA: MIT Press.

Greene, J. (2013). *Moral tribes: Emotion, reason and the gap between us and them*. London, UK: Atlantic Books Ltd.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111,* 364–371.

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers evidence for an innate moral core. *Current Directions in Psychological Science, 22,* 186–193. http://dx.doi.org/10.1177/0963721412470687

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106,* 766–794. http://dx.doi.org/10.1037/0033-295X.106.4.766

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92,* 1644–1655. http://dx.doi.org/10.1257/000282802762024700

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General, 133,* 23–30. http://dx.doi.org/10.1037/0096-3445.133.1.23

Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics, 114,* 14–18. http://dx.doi.org/10.1016/j.jpubeco.2013.11.006

Jordan, A. H., & Monin, B. (2008). From sucker to saint: Moralization in response to self-threat. *Psychological Science, 19,* 809–815. http://dx.doi.org/10.1111/j.1467-9280.2008.02161.x

Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of Personality and Social Psychology, 62,* 26–37. http://dx.doi.org/10.1037/0022-3514.62.1.26

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291. http://dx.doi.org/10.2307/1914185

Kirmani, A. (2015). Neatly tied with a bow. *Journal of Consumer Psychology, 25,* 185–186. http://dx.doi.org/10.1016/j.jcps.2015.02.002

Kraus, M. W., & Callaghan, B. (2016). Social class and prosocial behavior: The moderating role of public versus private contexts. *Social Psychological and Personality Science, 7,* 769–777.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin, 30,* 328–339.

Kruglanski, A. W., Bélanger, J. J., Chen, X., Köpetz, C., Pierro, A., & Mannetti, L. (2012). The energetics of motivated cognition: A force-field analysis. *Psychological Review, 119,* 1–20. http://dx.doi.org/10.1037/a0025488

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498. http://dx.doi.org/10.1037/0033-2909.108.3.480

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology, 25,* 272–285. http://dx.doi.org/10.1016/0022-1031(89)90023-1

Larrick, R. P. (1993). Motivational factors in decision theories: The role of self-protection. *Psychological Bulletin, 113,* 440–450. http://dx.doi.org/10.1037/0033-2909.113.3.440

Lin, S. C., Zlatev, J. J., & Miller, D. T. (2017). Moral traps: When self-serving attributions backfire in prosocial behavior. *Journal of Experimental Social Psychology, 70,* 198–203. http://dx.doi.org/10.1016/j.jesp.2016.11.004

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25,* 147–186.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45,* 633–644. http://dx.doi.org/10.1509/jmkr.45.6.633

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K., Tuller, D., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48,* 774–777. http://dx.doi.org/10.1016/j.jesp.2011.12.017

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4,* 344–357. http://dx.doi.org/10.1111/j.1751-9004.2010.00263.x

Miller, D. T., & Monin, B. (2016). Moral opportunities versus moral tests. In P. van Lange, J. Forgas, & L. Jussim (Eds.), *The social psychology of morality* (pp. 40–55). New York, NY: Psychology Press. http://dx.doi.org/10.4324/9781315644189-3

Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method*. Baltimore, MD: John Hopkins University Press.

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In D. Naraez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 341–354). New York, NY: Cambridge University Press.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81,* 33–43. http://dx.doi.org/10.1037/0022-3514.81.1.33

Nisan, M. (1990). Moral balance: A model of how people arrive at moral decisions. In T. E. Wren (Ed.), *The moral domain* (pp. 283–314). Cambridge, MA: MIT Press.

Paciello, M., Fida, R., Cerniglia, L., Tramontano, C., & Cole, E. (2013). High cost helping scenario: The role of empathy, prosocial reasoning and moral disengagement on helping behavior. *Personality and Individual Differences, 55,* 3–7. http://dx.doi.org/10.1016/j.paid.2012.11.004

Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization, 93,* 374–383. http://dx.doi.org/10.1016/j.jebo.2013.03.030

Polman, E. (2012). Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes, 119,* 141–150. http://dx.doi.org/10.1016/j.obhdp.2012.06.005

Rabin, M. (2000). Risk aversion and expected utility: A calibration theorem. *Econometrica, 68,* 1281–1292. http://dx.doi.org/10.1111/1468-0262.00158

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20,* 523–528. http://dx.doi.org/10.1111/j.1467-9280.2009.02326.x

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology, 53,* 396–403. http://dx.doi.org/10.1111/bjso.12060

Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115,* 181–190. http://dx.doi.org/10.1016/j.obhdp.2011.02.001

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science, 24,* 125–130. http://dx.doi.org/10.1177/0963721414553264

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review, 20,* 27–48. http://dx.doi.org/10.1177/1088868314568811

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445. http://dx.doi.org/10.1037/1082-989X.7.4.422

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science, 13,* 255–259.

Stone, E. R., Yates, A. J., & Caruthers, A. S. (2002). Risk taking in decision making for others versus the self. *Journal of Applied Social Psychology, 32,* 1797–1824. http://dx.doi.org/10.1111/j.1559-1816.2002.tb00260.x

Suls, J., Witenberg, S., & Gutkin, D. (1981). Evaluating reciprocal and nonreciprocal prosocial behavior: Developmental changes. *Personality and Social Psychology Bulletin, 7,* 25–31. http://dx.doi.org/10.1177/014616728171005

Tsang. (2002). Moral rationalization and the integration of situational factors and psychological processes in immoral behavior. *Review of General Psychology, 6,* 25–50.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59,* 251–278. http://dx.doi.org/10.1086/296365

White, K., Simpson, B., & Argo, J. J. (2014). The motivating role of dissociative out-groups in encouraging positive consumer behaviors. *Journal of Marketing Research, 51,* 433–447. http://dx.doi.org/10.1509/jmr.12.0335

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35,* 345–411. http://dx.doi.org/10.1016/S0065-2601(03)01006-2

Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology, 105,* 740–748. http://dx.doi.org/10.1037/a0032662

Wray, L. D., & Stone, E. R. (2005). The role of self-esteem and anxiety in decision making for self versus others in relationships. *Journal of Behavioral Decision Making, 18,* 125–144. http://dx.doi.org/10.1002/bdm.490

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120,* 202–214.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making, 12,* 93–106.

Received March 23, 2018
Revision received August 26, 2019
Accepted September 3, 2019 ■