Morally Motivated? People Use Self-Sacrifice as a Cue for Moral Character

Yochanan E. Bigman[1], Kurt Gray[2], and Shoham Choshen-Hillel[1]

[1] The Hebrew University Business School, and The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem

[2] Department of Psychology and Neuroscience, The University of North Carolina at Chapel Hill

We human beings are naturally inclined to characterize the morality of others. Whereas some actions lead us to immediate moral characterization of the agent, others can be more difficult to evaluate. Prior research suggests that when judging ambiguous actions, we rely upon our perception of the agent's motivation. Yet we are often not privy to the reasons behind agents' actions, making us search for cues of their moral motivation. Building on research on motivation attribution and person perception, we suggest here that self-sacrifice, the willingness to incur a personal cost, is a powerful cue that judges use to infer an agent's motivation. We hypothesize that people will be more likely to judge an agent positively when their action involves self-sacrifice, as it is perceived as a reflection of moral motivation. However, when the agent's motivation is clearly immoral, self-sacrifice will not affect moral judgment. Six vignette studies (N = 3,931) each manipulating the actor's self-sacrifice, supported this hypothesis across various domains, including views of political policies, assessment of risky military decisions, and monetary decisions.

Keywords: moral judgment, motivation attribution, self-interest, self-sacrifice, person perception

People use various cues to infer the morality of others. One obvious cue to a person's morality is the outcome of their action (Uhlmann et al., 2014). For example, a person saving someone else's life will most likely be judged as moral. Yet the morality of some actions is difficult to judge, especially if they involve both positive and negative outcomes. For example, how might people judge the morality of someone who saved one life while taking another?

It has been suggested that motivation is key to understanding a person's moral character (Carlson et al., 2022). Unfortunately, motivation is often obscured and must be inferred. Consider an army

officer who sends their soldiers on a highly risky mission that could prevent a terrorist attack but will likely result in casualties. They might be trying to save civilians but might also not care enough about their soldiers and may be trying to get promoted.

Here, we argue that a powerful cue to inferring moral motivation is *self-sacrifice*—whether a person was knowingly willing to accept a personal cost. We suggest that, when an action is morally ambiguous, if the agent was willing to pay a personal price, observers are more likely to infer that the agent's motivation was moral, and hence categorize them as morally good. For example, we propose that people are more likely to infer that the officer who sent his soldiers on a risky mission is motivated by a moral motivation if he joined them himself than if he did not. This difference in motivation attribution will cause people to judge the self-sacrificing officer more positively than the officer who did not self-sacrifice, even if the outcomes of their actions are the same (success of the mission and number of soldiers killed). Thus, in this article, we test whether the presence of *self-sacrifice* affects the motivation attributed to an agent and consequently the way the agent is morally judged.

Motivation Matters for Morality

Research shows that people care about the moral character of the agent. According to the "person-centered approach" to morality, people care not only about the morality of actions but about what those actions mean about the *agents* performing them (Goodwin et al., 2014; Uhlmann et al., 2015). The person-centered approach argues that morality is about people navigating the social world, predicting the behavior of others and avoiding those who have a large potential for harm (Chakroff et al., 2017; McHugh et al., 2022).

One way in which people might form judgments about the moral character of others is by evaluating the motivation underlying their behavior—why they behaved the way they did (Bigman & Tamir, 2016; Carlson et al., 2022; Choshen-Hillel et al., 2020; Reeder et al., 2002). For example, people generally view liars as immoral, but if they lie for a prosocial motive, their moral character is seen more positively, and they are trusted even more than those who told the truth (Levine & Schweitzer, 2014, 2015). Another example for the importance of

motivation for moral judgment comes from a study that shows that people judge those who merely hope for a natural disaster to happen as immoral, as this motivation is seen as reflecting an immoral character (Inbar et al., 2012).

Motivation refers to the reason *why* a person behaved the way they did (Bargh et al., 2010). Note that motivation is not the same as intention, which refers to whether an action and its outcome were planned (Carlson et al., 2022). Both perceived intention and perceived motivation affect moral judgment. Perceived intention affects moral judgment, as intentional harm—an action that was believed to cause harm—is judged as worse than unintentional harm (Alicke, 2000; Cushman, 2008). Motivation adds information above and beyond intention. For example, both a veterinarian and a psychopath may intentionally kill cats. But since their motivations are perceived as different, this leads to starkly different moral evaluations for the same act.

When an action has a clear positive outcome (helping others) or a clear negative outcome (harming others), people are likely to infer that the agent's motivation is consistent with the outcome. However, sometimes, an action is morally ambiguous—it has both morally positive and morally negative consequences—which makes it challenging to infer the true motive behind an act. A politician might oppose affirmative action because they are bigoted, or because they believe affirmative action is inefficient. A military officer might order a risky attack because they think it is important for area security, but they might make the same decision to increase their likelihood of being promoted. In such cases of moral ambiguity, how do people infer the motives of others?

Self-Sacrifice

We propose that self-sacrifice serves as a powerful cue for inferring whether an agent's motivation is moral or not. Previous research indicates that when an action benefits the agent, observers often attribute selfish motives to the agent, leading to negative moral judgments (Carlson & Zaki, 2018; Miller, 1999; Miller & Ratner, 1998; Ratner & Miller, 2001). However, the opposite is not necessarily true for self-sacrifice. Positive behaviors (e.g., helping the environment) and negative behaviors (e.g., harming the environment) are not always judged in a symmetrical manner (Klein & Epley, 2014; Knobe, 2003; Pizarro et al., 2003).

For example, people are more inclined to perceive harmful actions as intentional compared to actions that result in benefit (Knobe, 2003). Given this potential asymmetry, the dynamics of negative moral judgments should be tested separately. In this investigation, we focus on self-sacrifice and propose that it reduces attributions of selfish motivation and increases attributions of prosocial intent, thereby enhancing moral judgments.

We define self-sacrifice as a voluntary action that carries the potential for harm to the actor. This harm can be financial, physical, or emotional. Harm to the self need not be realized to affect moral judgment; it should be enough to know that the agent was willing to make a sacrifice. On the other hand, an action that an agent takes without expecting personal harm—even if it results in harm—should not be seen as self-sacrifice, as it carries no information about the agent's motivation.

We argue that in cases where an action is morally ambiguous, when people view an agent who self-sacrifices, they infer that she/he is less motivated by selfishness and more motivated by care for others, which is seen as a more moral motivation. We expect that self-sacrifice would serve as a particularly strong cue to motivation, because it defies the default assumption that others are mainly driven by self-interest (Miller, 1999; Newman & Cain, 2014; Ratner & Miller, 2001). The moral motivation people associate with self-sacrificing agent will lead them to judge the agent as more moral.

The Current Research

In six studies, we tested the hypothesis that people use self-sacrifice as a cue to agents' moral motivation, which in turn affects their moral judgment. Figure 1 shows this core theoretical

Figure 1
Theoretical Model
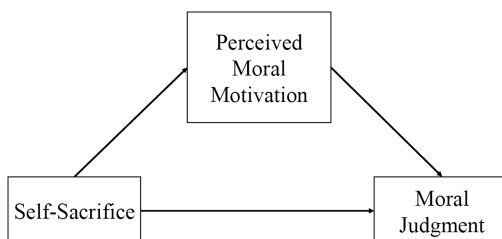
Note. The model predicts that an agent's self-sacrifice affects how others morally judge them. This effect is driven by an altered perception of the agent's moral motivation.

model. Study 1 tested this model by examining the effect of an agent's self-sacrifice on perceived motivation and moral judgments in two scenarios: a military commander ordering a risky mission (Study 1a) and a person opposing affirmative action (Study 1b). Study 2 examined whether the self-sacrifice has a long-lasting effect on moral judgment, by shaping people's perception of the agent's moral motivation in a subsequent action. Studies 3a and 3b further established the role of perceived motivation in the effect of self-sacrifice. They test the idea that self-sacrifice affects moral judgment only when the agent's motivation is ambiguous and does not affect moral judgment when the agent's motivation is clearly immoral. Study 4 disentangled the impact of two parts of actions involving self-sacrifice that often co-occur: the decision to self-sacrifice and the outcome of sacrifice. According to our hypothesis, making a self-sacrificial decision should suffice to impact judgment, as it signals motivation, even if the action did not result in a sacrificial outcome.

Finally, judging self-sacrificing agents more positively should manifest itself in people's behavior toward them. Study 5 tested this idea by examining whether people would donate more to a cause when the fundraisers were fundraising for a cause that may personally harm them. Study 6 tested this by examining whether a politician's self-sacrifice (support of a policy that personally harms him) affects how people judge him and their willingness to vote for him.

Ethics and Material Sharing

All studies were approved by the institutional review board of the Hebrew University of Jerusalem and of the University of North Carolina at Chapel Hill. All studies were preregistered (links appear in the method sections below). Participants were all 18 and older. Study materials, anonymized data, and additional online material are available at https://osf.io/5rdxu/?view_only=b4e2828074a f40c68c0adc75529bbf7a. All measures, manipulations, and exclusions are reported. Preregistered sample sizes were determined by a priori power analyses reported in the method sections.

Study 1a: Commander Ordering a Risky Attack

In this study, we tested our main hypothesis (outlined in Figure 1), whereby an agent who

self-sacrificed is judged more positively morally, because their motivation is perceived as more moral. We manipulated self-sacrifice by describing a U.S. Army commander who joined a risky attack and was killed in action or did not join the attack. We predicted that participants would judge the commander and his decision to send the soldiers on the risky mission as more moral when he had joined them himself—even though the overall outcome was worse in that case, with a higher number of causalities (the soldiers plus the commander). We further predicted that more positive judgments of the commander in the self-sacrifice condition would be mediated by greater attributions of moral motivation.

Method

Preregistration

Participants

Two hundred and forty-one participants (45.6% male, 54.4% female; age: $M = 34.3$, $SD = 10.9$) from the United States completed the study on Amazon's Mechanical Turk in exchange for 40 cents. We aimed for a sample size of 120 participants per cell in Studies 1a and 1b. Accounting for participants who might fail attention checks and therefore be excluded from the analysis, this sample size provides a power of 0.95 to detect a medium effect size (Cohen's $d = 0.5$, calculated with G*Power 3.1.9.2). As specified in the preregistration, we excluded from the analysis participants who failed to answer correctly at least one of the two attention checks, leading to the exclusion of 35 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.39$.

Procedure

All participants read the following vignette:

Jim Davis is a Colonel in the U.S. Army deployed in Afghanistan. A local Al-Qaeda terrorist cell has been attacking U.S. Army supply lines. If these terrorists are not dealt with, they will eventually carry out a terror attack killing some American soldiers.

Colonel Davis assesses the situation and concludes that a local U.S. Army squad located nearby can take the Al-Qaeda cell out, but such an attack will be very risky and many soldiers may die.

It is unclear if eliminating the terrorist group is worth the risk to the soldiers.

Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack.

Participants were randomly assigned to a no self-sacrifice or to a self-sacrifice condition. Participants in the no self-sacrifice condition read the following ending: "The squad successfully kills the terrorists. Ten of the 12 squad members are killed during the battle." Participants in the self-sacrifice condition instead read a different ending: "Colonel Davis personally joins the squad in their mission. The squad successfully kills the terrorists. Colonel Davis and 10 of the 12 squad members are killed during the battle."

Assessing Moral Judgment.   After reading the scenario, participants in both conditions reported their moral judgment of Colonel Davis by answering the following six questions adapted from previous research (Ames & Fiske, 2013; Critcher et al., 2013; Cushman, 2008; Pizarro et al., 2003): "How moral or immoral is Colonel Davis' decision to order the attack?" (1 = *extremely immoral*; 9 = *extremely moral*), "How much blame or praise should Colonel Davis receive for his decision to order the attack?" (1 = *extreme blame*; 9 = *extreme praise*), "To what extent would you describe Colonel Davis as a good or a bad person?" (1 = *bad person*; 9 = *good person*), "What punishment or reward should Colonel Davis receive for his decision to order the attack?" (1 = *harsh punishment*; 9 = *great reward*),[1] "How right or wrong was Colonel Davis' decision to order the attack?" (1 = *very wrong*; 9 = *very right*), and "Colonel Davis' decision was" (1 = *completely forbidden*; 9 = *completely permissible*). We created a composite moral judgment score by averaging the six items, Cronbach's $\alpha = .92$.

Assessing Perceived Moral Motivation.   After rating their moral judgment of Colonel Davis, participants reported their perception of Colonel Davis' moral motivation by rating their agreement

---

[1] Note that because in the self-sacrifice condition Colonel Davis dies, some participant might have found this item strange. Results for Study 1a and its replication remain unchanged when excluding this item from the analysis.

A per-item analysis for the effect of self-sacrifice on all moral judgment items in all studies can be found in Appendix A in the projects' open science framework [https://osf.io/5rdxu/?view_only=b4e2828074af40c68c0adc75529bbf7a]. Most of the items were significant also individually. We clearly mark which items in which studies were not.

with the following statements: "Colonel Davis wanted to save as many soldiers as possible," "Colonel Davis wanted to maximize the total number of soldiers saved," and "Colonel Davis wanted to minimize the number of soldiers who die." Each item was rated on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We created a composite index of moral motivation by averaging participants' ratings of these three items (Cronbach's $\alpha = .90$).

Participants then answered two attention checks. The first about whether they read about a U.S. commander or an Al-Qaeda commander, and the second about whether the commander joined the attack or not. Finally, participants provided demographic information.

Results

As predicted, an independent samples *t* test revealed that participants judged Colonel Davis as more moral in the self-sacrifice condition ($M = 6.11$, $SD = 1.61$) than in the no self-sacrifice condition, $M = 5.17$, $SD = 1.33$, $t(204) = 4.59$, $p < .001$, Cohen's $d = 0.64$. An independent samples *t* test revealed that participants attributed more moral motivation to Colonel Davis in the self-sacrifice condition ($M = 5.12$, $SD = 1.35$) than in the no self-sacrifice condition, $M = 4.67$, $SD = 1.47$, $t(204) = 2.31$, $p = .022$, Cohen's $d = 0.32$. This difference suggests that participants saw Colonel Davis' self-sacrifice as evidence that he was more motivated by sincere care for his troops.

To test whether perceived moral motivation mediated the link between self-sacrifice and moral judgment, we performed a bootstrapping mediation analysis (we used Preacher & Hayes, 2008; 5,000 iterations, Model 4 for all mediation analyses in this article), coding the self-sacrifice condition as 1 and the no self-sacrifice condition as −1. As predicted, the direct positive effect of self-sacrifice on moral judgment ($b = 0.47$, $SE = 0.10$, $p < .001$) was partially mediated by an indirect effect of attribution of moral motivation ($b = 0.12$, $SE = 0.06$, 95% CI [0.02, 0.24]). When accounting for the mediation by attribution of moral motivation, the direct effect of self-sacrifice on moral judgment was still significant ($b = 0.35$, $SE = 0.09$, 95% CI [0.17, 0.53]), see Figure 2. These results are consistent with our theory that the increase in positive moral judgment caused by self-sacrifice is driven, at least partially, by increased attributions of being motivated by a care for the welfare of others.

Figure 2
Mediation Analysis in Study 1a Reveals That Perceived Moral Motivation Statistically Mediates the Effect of Self-Sacrifice on Moral Judgment
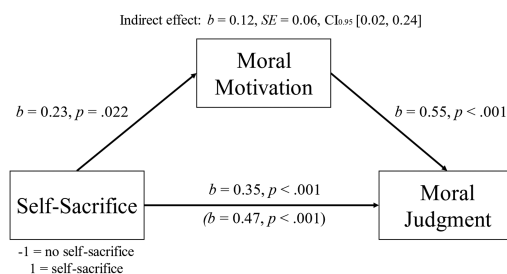
Note. SE = standard error; CI = confidence interval.

Replication With a Sample of Israeli College Students

To examine the generalizability of the effect of self-sacrifice on moral judgment, we replicated Study 1a with a different sample: college students at an Israeli university (preregistered at: https://aspredicted.org/u92nb.pdf, $N = 198$, 41.4% male, 58.1% female, 0.5% did not report their gender; age: $M = 24.9$, $SD = 6.4$; after 42 preregistered exclusions, with a power of 0.8 to detect an effect size of Cohen's $d = 0.40$). We used a similar scenario describing a commander who self-sacrificed by joining a military operation (vs. not), adapting the scenario to the context of the Israel–Hezbollah conflict (the study was conducted in October 2018). Participants read about an Israeli commander ordering a risky operation against Hezbollah terrorists. The commander either joined the mission himself or did not. Here, we measured only moral judgment. Replicating our main finding from Study 1a, we found that the commander was judged more positively in the self-sacrifice condition ($M = 6.50$, $SD = 1.21$) than in the no self-sacrifice condition, $M = 6.00$, $SD = 1.14$, $t(196) = 2.96$, $p = .003$, Cohen's $d = 0.42$. See additional online material at https://osf.io/5rdxu/?view_only=b4e2828074af40c68c0adc75529bbf7a for full study materials and results.

Discussion

Study 1a is consistent with our theoretical model. Self-sacrifice led to more positive moral judgments. Importantly, the effect of self-sacrifice on moral judgment is consistent with mediation by

increased attributions of moral motivation. When the commander himself joined the risky mission, participants attributed to him a more moral motivation, which was associated with judging him more positively than when he had not joined the mission. Note that the commander's decision was viewed more positively in the self-sacrifice condition, although it involved the same questionable behavior (i.e., risking the lives of the soldiers) and led to objectively worse outcomes (i.e., his own death *in addition* to that of 10 soldiers). However, it is possible that in Study 1a, participants judged the commander more positively when he joined his soldiers not because he sacrificed but because he exhibited courage and because joining risky missions might have been seen as part of his duty. We address this limitation in Study 1b.

Study 1b: Opposing Affirmative Action

In Study 1b, we aimed to extend the findings from Study 1a by testing the same hypothesis in a different context, where perceptions of courage and duty were not associated with self-sacrifice. Here, participants read about a student who opposed gender-equality affirmative action in admissions to engineering graduate school. We manipulated self-sacrifice by informing participants that the student was either a female college student who would have a lower chance of acceptance if affirmative action was canceled or a male who would not be personally affected. We measured participants' moral judgment of the student's opposition to affirmative action, and the moral motivation they attributed to them. We predicted that participants would view the student opposing gender-based affirmative action as more moral if she made a personal sacrifice by her opposition. We also predicted that the effect of self-sacrifice on moral judgment would be mediated by a more moral motivation attributed to the sacrificing student.

Method

Preregistration

https://aspredicted.org/hr7s-nsrc.pdf.

Participants

Two hundred and forty participants from the United States (48.8% male, 48.3% female; 2.9%

other or preferred not to disclose; age: $M = 33.5$, $SD = 12.1$) completed the study on Prolific in exchange for 40 cents. As specified in the preregistration, we excluded from the analysis participants who failed at least one of the three attention checks, leading to the exclusion of 12 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.37$.

Procedure

To test participants' attention, they were first asked to indicate the day of the week and describe what they had for breakfast. They were then randomly assigned to a no self-sacrifice or a self-sacrifice condition. Participants in the no self-sacrifice condition read the following:

> John Wilson is an undergraduate college student who opposes gender-equality affirmative action in admission to graduate degrees in engineering. In an interview he said:

> Affirmative action is a bad idea. We need the best engineers we can get. I wouldn't want to cross a bridge that was built by an engineer that got into engineering school because of affirmative action.

In the self-sacrifice condition, participants read that the student was Amelia Wilson, a woman who is applying to graduate school in engineering, and therefore might pay a personal price if there is no affirmative action:

> Amelia Wilson is an undergraduate college student who is applying to a graduate degree in engineering. She opposes gender-equality affirmative action in admission to graduate degrees in engineering. In an interview she said:

> I know affirmative action might help me get into the program I want. But still, I think that affirmative action is a bad idea. We need the best engineers we can get. I wouldn't want to cross a bridge that was built by an engineer that got into engineering school because of affirmative action.

Assessing Perceived Self-Sacrifice. After participants read the vignette, they were asked to assess the agent's (John or Amelia Wilson) self-sacrifice in his/her opposition to affirmative action. Specifically, they had to rate their agreement with the following three statements on a 1 (*strongly disagree*) to 9 (*strongly agree*) scale: "John/Amelia Wilson sacrifices his/her personal interest by opposing affirmative action," "John/Amelia Wilson would personally be harmed if affirmative action was stopped" and "John/Amelia

Wilson's opposition to affirmative action is against his/her self-interest." We created a composite index of perceived self-sacrifice by averaging participants' ratings of these three items (Cronbach's α = .79).

Assessing Perceived Moral Motivation. Participants were then asked to rate their agreement with the following three statements on a 1 (*strongly disagree*) to 9 (*strongly agree*) scale: "John/Amelia Wilson cares about engineering quality," "John/Amelia Wilson cares about safety," and "John/Amelia Wilson wants engineering schools to train the best possible candidates." We created a composite index of moral motivation by averaging participants' ratings of these three items (Cronbach's α = .94).

Assessing Moral Judgment. Participants were asked to answer four questions about their moral judgment of the agent's position on affirmative action: "How moral or immoral is John/Amelia Wilson's position on affirmative action?" (1 = *extremely immoral*; 9 = *extremely moral*), "How much blame or praise should John/Amelia Wilson receive for his/her position on affirmative action?" (1 = *extreme blame*; 9 = *extreme praise*), "To what extent would you describe John/Amelia Wilson as a good or a bad person?" (1 = *bad person*; 9 = *good person*), and "How right or wrong was John/Amelia Wilson's position on affirmative action?" (1 = *very wrong*; 9 = *very right*). These four questions were identical to four of the six questions we used in Study 1a. We did not include questions about the permissibility of the action or about the punishment and reward for this study (questions that we used in Study 1a), because these items were less applicable for judging an ideological position such as supporting or opposing affirmative action. We created a composite moral judgment score by averaging participants' ratings of the four items (Cronbach's α = .95).

Participants then answered two attention checks. The first asked whether the student they read about was named John Wilson or Amelia Wilson and the second whether this student supported or opposed gender-equality affirmative action. Finally, participants provided demographic information.

Results

Confirming the self-sacrifice manipulation, participants perceived higher levels of self-sacrifice in the self-sacrifice condition ($M = 5.67$,

$SD = 1.79$) than in the no self-sacrifice condition, $M = 3.18$, $SD = 1.87$, $t(226) = 10.22$, $p < .001$, Cohen's $d = 1.36$.

In line with our main prediction, participants judged the student as more moral in the self-sacrifice condition ($M = 6.08$, $SD = 1.99$) than in the no self-sacrifice condition ($M = 4.11$, $SD = 2.11$), $t(225) = 7.27$, $p < .001$, Cohen's $d = 0.96$.[2] Furthermore, participants attributed more moral motivation to the student in the self-sacrifice condition ($M = 8.04$, $SD = 1.38$) than in the no self-sacrifice condition, $M = 6.12$, $SD = 2.45$, $t(226) = 7.22$, $p < .001$, Cohen's $d = 0.97$. To test whether perceived moral motivation mediated the link between self-sacrifice and moral judgment, we performed a bootstrapping mediation analysis, coding the self-sacrifice condition as 1 and the no self-sacrifice condition as −1. As predicted, the total positive effect of self-sacrifice on moral judgment ($b = 0.99$, $SE = 0.14$, $p < .001$) was partially mediated by an indirect effect of attribution of moral motivation ($b = 0.62$, $SE = 0.09$, 95% CI [0.45, 0.80]). When accounting for the mediation by attribution of moral motivation, the direct effect of self-sacrifice on moral judgment was still significant ($b = 0.37$, $SE = 0.12$, 95% CI [0.14, 0.60]). This is consistent with our theory that the increase in positive moral judgment caused by self-sacrifice is driven, at least partially, by increased attributions of being morally motivated by a care for quality engineering.

To further test the validity of our model and examine whether *perceived* (rather than manipulated) self-sacrifice affected moral judgment by affecting perceived moral motivation, we conducted another mediation analysis. In this analysis, we used perceived self-sacrifice, rather than experimental condition, as the independent variable. We included perceived moral motivation as the mediator and moral judgment as the dependent variable. The analysis revealed a main effect for perceived self-sacrifice on moral judgment ($b = 0.27$, $SE = 0.07$, 95% CI [0.14, 0.40]) that was partially mediated by an indirect effect of perceived moral motivation ($b = 0.18$, $SE = 0.05$, 95% CI [0.07, 0.26]) consistent with our model.

---

[2] We tested whether this effect interacted with Gender with an exploratory 2 (Condition: self-sacrifice, no self-sacrifice) × 2 (Gender: male, female) between subjects ANOVA. We found a significant condition effect, $F(1, 218) = 50.12$, $p < .001$, partial $\eta^2 = .187$. The effect for gender and the Gender × Condition interaction did not reach significance, ps = .062 (partial $\eta^2 = .016$) and .107 (partial $\eta^2 = .012$), respectively.

Discussion

Study 1b provided further support for our hypothesis. We replicated the main finding from Study 1a in a new context, with a less extreme case of self-sacrifice, where courage and duty could not account for the effect. As in Study 1a, self-sacrifice led to more positive moral judgments. Importantly, the effect of self-sacrifice on moral judgment was consistent with mediation, though partially, by greater perceived moral motivation. When the protagonist was likely to be negatively affected by the policy, participants perceived her motivation as more moral, which was associated with a more positive judgment than a protagonist who would not be negatively affected.

Study 2: Past Sacrifice

In Studies 1a and 1b, we found that self-sacrifice led to more positive moral judgment. In both these studies, the agent is judged on the same action they have self-sacrificed: by joining a risky attack in Study 1a or opposing personally helpful affirmative action in Study 1b. According to our hypothesis, self-sacrifice is seen as reflecting moral motivation, which is closely related to moral character (Carlson et al., 2022). Moral character is often seen a disposition, transcending specific acts to describe the general disposition of a person (Goodwin et al., 2014; Pizarro & Tannenbaum, 2012; Uhlmann et al., 2015). Therefore, the effect of self-sacrifice in a specific event might extend to how an agent is judged in unrelated, later events.

To test this possibility, in Study 2, participants read about a military commander who either decided to self-sacrifice in the past or did not. Participants were then asked to judge a separate, later decision made by the commander, where he had ordered his soldiers to go on a risky attack, and to assess his motivation. We predicted that the effect of self-sacrifice would extend to future behaviors, and therefore that the commander would be judged more positively when he self-sacrificed in the past (vs. not).

Method

Preregistration

https://aspredicted.org/ZPN_MFV.

Participants

Three hundred and sixty-eight participants from the United States (50% male, 48.6% female, 1.4% other or preferred not to disclose; age: $M = 35.8$, $SD = 14.2$) completed the study on Prolific in exchange for 50 cents. We aimed for a sample size of 180 participants per cell. Taking into account participants who might fail attention checks and therefore be excluded from the analysis, this sample size provides a power of 0.95 to detect a medium-small effect size (Cohen's $d = 0.4$, calculated with G*Power 3.1.9.2). As specified in the preregistration, we excluded from the analysis participants who failed to answer correctly at least one of the three attention checks, leading to the exclusion of 16 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.30$.

Procedure

Participants first answered an attention check asking them what day of the week it was and what they had for breakfast. We excluded participants whose answers did not include both components (day and meal) or that were nonsensical (e.g., providing their prolific ID number). Participants were then randomly assigned to one of two conditions. In the no self-sacrifice condition, participants instead read the following:

> Jim Davis is a Colonel in the U.S. Army deployed in Iraq. A few years ago, he was serving in Afghanistan. His soldiers went on a dangerous mission, and Colonel Davis was supervising them from the base. The soldiers' car was hit by terrorists and one of the soldiers was badly injured.

In the self-sacrifice condition, participants read the following:

> Jim Davis is a Colonel in the U.S. Army deployed in Iraq. A few years ago, he was serving in Afghanistan. His soldiers went on a dangerous mission, and Colonel Davis decided to join them, although he could have stayed at the base. Colonel Davis' car was hit by terrorists and he was badly injured.

All participants then answered an attention check, on whether Colonel Davis joined his soldiers on a dangerous mission (options: yes, no, not mentioned) and read the following:

> Recently, a local terrorist cell has been attacking U.S. Army supply lines. If these terrorists are not dealt with, they will eventually carry out a terror attack killing some American soldiers.

Colonel Davis assesses the situation and concludes that a local U.S. Army squad located nearby can take the terrorist cell out, but such an attack will be very risky and many soldiers may die.

It is unclear if eliminating the terrorist group is worth the risk to the soldiers. In such cases, some officers decide to order the attack. Other officers see it as too risky and do not order the attack.

Despite the risk to the soldiers, Colonel Davis orders the soldiers to carry out the attack.

The terrorists were killed. Ten of the twelve squad members were also killed during the battle.

Participants were asked to rate how moral they thought the motivation of Colonel Davis was in ordering the attack on the terrorists (using the same items as in Study 1a, Cronbach's $\alpha = .94$) and their moral judgment of Colonel Davis (using the same items as in Study 1a, Cronbach's $\alpha = .89$). Participants answered another attention check whether Colonel Davis died or not. Finally, they provided basic demographic information.

Results

As predicted, an independent samples t test revealed that participants judged Colonel Davis as more moral in the self-sacrifice condition ($M = 5.27$, $SD = 1.37$) than in the no self-sacrifice condition, $M = 4.73$, $SD = 1.43$, $t(350) = 3.59$, $p < .001$, Cohen's $d = 0.39$. An independent samples t test also revealed that participants attributed more moral motivation to Colonel Davis in the self-sacrifice condition ($M = 4.20$, $SD = 1.69$) than in the no self-sacrifice condition, $M = 3.63$, $SD = 1.64$, $t(350) = 3.23$, $p < .001$, Cohen's $d = 0.34$.

To test whether perceived moral motivation mediated the link between self-sacrifice and moral judgment, we performed a bootstrapping mediation analysis, coding the self-sacrifice condition as 1 and the no self-sacrifice condition as -1. As predicted, the direct positive effect of self-sacrifice on moral judgment ($b = 0.27$, $SE = 0.07$, $p < .001$) was partially mediated by an indirect effect of attribution of moral motivation ($b = 0.14$, $SE = 0.05$, 95% CI [0.05, 0.24]). When accounting for the mediation by attribution of moral motivation, the direct effect of self-sacrifice on moral judgment was still significant ($b = 0.13$, $SE = 0.06$, 95% CI [0.01, 0.25]). Plausibly, the increase in positive moral judgment caused by self-sacrifice is partially driven by increased attributions of being motivated by a care for the welfare of others.

Discussion

The results of Study 2 extend our previous findings. We find that the effect of self-sacrifice is not limited to actions that directly involve the self-sacrifice. Instead, self-sacrifice in the past is used as a clue for inferring motivation of future actions, and thereby leads to more positive moral judgment. In particular, in Study 2, the commander who self-sacrificed in the past was judged more positively a few years after the self-sacrifice had happened, compared to the commander who did not self-sacrifice. Note that in Study 2, the decision to self-sacrifice (i.e., the colonel joining his troops) *precedes* the questionable moral decision (i.e., risking the soldiers). In Study 1a, the order is reversed, such that the decision to self-sacrifice *follows* the questionable moral decision. The fact that the self-sacrifice effect is observed in both studies suggests that it might be robust to the temporal order of the decisions.

Studies 1 and 2 revealed that self-sacrifice elevated moral judgment. A mediation analysis further showed that the self-sacrifice effect might be partially explained by greater attribution of moral motivation to a self-sacrificing agent. However, mediation analyses only provide partial support for an inference about the causal role of the mediating variable (Rohrer, 2024). In Studies 3a and 3b, we manipulate moral motivation directly.

Study 3a: Moderation by Moral Motivation

According to our theory, self-sacrifice affects moral judgment via changing how people perceive an agent's moral motivation. If an agent's motivation is clearly immoral, however, then self-sacrifice should not affect moral judgment, because it cannot make the motivation seem more moral. In Study 3a, we therefore tested the causal role of moral motivation in the effect of self-sacrifice on moral judgment. If our hypothesis is correct, then this is also a boundary condition for the self-sacrifice effect. We used the commander scenario from Study 1a and manipulated self-sacrifice and moral motivation orthogonally. We added two new conditions where the agent's motivation was clearly immoral: an Al-Qaeda leader ordering an attack on U.S. Army soldiers,

with or without self-sacrifice. We predicted that self-sacrifice will boost moral judgment when the decision is morally ambiguous (an American commander ordering an attack that might risk his soldiers), but not when the motivation is clearly immoral (an Al-Qaeda commander ordering an attack on American soldiers, that might risk the terrorists).

Method

Preregistration

https://aspredicted.org/t2yc-dvgk.pdf.

Participants

Four hundred and eighty-four participants from the United States (45.7% male, 54.3% female; age: $M = 36.5$, $SD = 11.3$) completed the study on MTurk in exchange for 40 cents. We aimed for a sample size of 120 participants per cell. Taking into account participants who might fail attention checks and therefore be excluded from the analysis, this sample size provides a power of 0.95 to detect a medium-small effect size (Cohen's $d = 0.35$, calculated with G*Power 3.1.9.2). As specified in the preregistration, we excluded from the analysis participants who failed to answer correctly at least one of the three attention checks, leading to the exclusion of 61 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.27$. We note that this study was conducted in September 2018.

Procedure

Participants were randomly assigned to one of four conditions, according to a 2 (Self-Sacrifice: Yes, No) × 2 (Agent's Morality: Moral (U.S. Army), Immoral (Al-Qaeda)) between-subject design. The moral agent conditions were very similar to those used in Study 1a, describing the U.S. Army Colonel Davis who carried out a risky attack against Al-Qaeda. In the self-sacrifice condition participants were told that Colonel Davis joined the attack and died. In the no self-sacrifice condition, participants were told that Colonel Davis did not join the attack. In the immoral conditions, participants instead read about a terrorist, Akhtar Mansur, a regional Al-

Qaeda leader in Afghanistan, who ordered an attack on a U.S. army squad:

> Akhtar Mansur is a regional Al-Qaeda leader in Afghanistan. A local U.S. Army squad has been attacking Al-Qaeda supply lines. If these U.S. soldiers are not dealt with, they will eventually carry out an attack killing some Al-Qaeda militants.
>
> Akhtar Mansur assesses the situation and concludes that a local Al-Qaeda cell located nearby can take the U.S. Army squad out, but such an attack will be very risky and many Al-Qaeda militants may die.
>
> It is unclear if eliminating the U.S. Army squad is worth the risk to the local Al-Qaeda cell.
>
> Despite the risk to the militants, Akhtar Mansur orders the local Al-Qaeda cell to carry out the attack.

Participants in the no self-sacrifice condition read the following ending:

> Akhtar Mansur does not join the Al-Qaeda cell in their mission. The Al-Qaeda cell successfully kills the U.S. Army soldiers. Ten of the twelve cell members are killed during the battle.

Participants in the self-sacrifice condition read a different ending:

> Akhtar Mansur personally joins the Al-Qaeda cell in their mission. The Al-Qaeda cell successfully kills the U.S. Army soldiers. Akhtar Mansur and ten of the twelve cell members are killed during the battle.

After reading the story, participants rated their moral judgment of the agent (Davis or Mansur, depending on the condition) with the same six items as in Study 1a (Cronbach's α = .96).[3]

Manipulation and Attention Checks

To ensure that the participants viewed the actions of the U.S. Army as more moral than those of Al-Qaeda, we asked participants to rate the morality of the action on a 1 (extremely immoral) to 9 (extremely moral) scale. Participants in the moral conditions were asked about the morality of the U.S. Army squad eliminating the Al-Qaeda cell and participants in the immoral conditions were asked about the morality of the Al-Qaeda cell eliminating the U.S. Army squad.

Participants then answered three attention checks: the two used in Study 1a, as well as a new one. In the new attention check, participants viewed a list of animals and countries and were

---

[3] We did not measure perceived motivation in this study.

asked to select all the animals. Finally, participants provided demographic information.

Results

Manipulation Check

Confirming the morality manipulation, participants rated the action of a U.S. Army squad eliminating an Al-Qaeda cell as more moral ($M = 6.53$, $SD = 1.83$) than an Al-Qaeda cell eliminating a U.S. Army squad ($M = 2.32$, $SD = 1.65$), $t(421) = 24.77$, $p < .001$, Cohen's $d = 2.42$.

Moral Judgment

A 2 (Self-Sacrifice: Yes, No) × 2 (Agent's Morality: Moral (U.S. Army), Immoral (Al-Qaeda)) between-subject analysis of variance (ANOVA) revealed a main effect for Agent's morality, $F(1, 419) = 370.99$, $p < .001$, $\eta_p^2 = .470$, whereby the U.S. Army leader was judged as more moral ($M = 5.07$, $SD = 1.74$) than the Al-Qaeda leader ($M = 2.19$, $SD = 1.47$). Replicating our previous findings, self-sacrifice had a significant effect on moral judgment, $F(1, 419) = 25.91$, $p < .001$, $\eta_p^2 = .058$, whereby the agents were judged as more moral in the self-sacrifice condition ($M = 4.09$, $SD = 2.44$) than in the no self-sacrifice condition ($M = 3.25$, $SD = 1.78$). Importantly, this main effect for self-sacrifice was qualified by a significant Self-Sacrifice × Agent's Morality interaction, $F(1, 419) = 18.61$, $p < .001$, $\eta_p^2 = .043$. Pairwise comparisons revealed that the U.S. Army officer was judged as more moral in the self-sacrifice condition ($M = 5.78$, $SD = 1.75$) than in the no self-sacrifice condition ($M = 4.37$, $SD = 1.44$), $F(1, 419) = 45.27$, $p < .001$, $\eta_p^2 = .097$, whereas there was no significant difference in how the Al-Qaeda leader was judged in the self-sacrifice ($M = 2.25$, $SD = 1.62$) and the no self-sacrifice ($M = 2.13$, $SD = 1.33$) conditions, $F(1, 419) = 0.29$, $p = .588$, $\eta_p^2 = .001$.

Discussion

The results of Study 3a demonstrate an important boundary condition for the effect of self-sacrifice on moral judgment: When an agent is clearly immoral—such as in the case of an Al-Qaeda terrorist—self-sacrifice does not improve moral judgment. It also supports the causal role of perceived moral motivation. When the motivation of the agent is clearly immoral—as in the case of a terrorist leader—self-sacrifice does not affect moral judgment. However, this study did not measure perceived moral motivation, so there is no direct evidence for our theory that self-sacrifice did not affect the judgment of the terrorist leader, because his motivation was perceived as clearly immoral. We added a measurement of perceived moral motivation in Study 3b.

Study 3b: Moderation by Moral Motivation—Measuring Perceived Motivation

In Study 3b, we aimed to further establish the causal role of perceived motivation in the effect of self-sacrifice on moral judgment. To do so, we used a similar method to that of Study 3a and added a measure of perceived moral motivation. This would allow us to directly examine the underlying role of perceived motivation. According to our theory, when the agent's motivation is clearly immoral, such as in the case of an Al-Qaeda terrorist, self-sacrifice should not affect moral judgment (because it cannot change the perception of moral motivation). We also expanded the diversity of our samples by recruiting university students from Israel as participants (rather than an online American sample).

Method

Preregistration

https://aspredicted.org/p5sv-s6bd.pdf.

Participants

As specified in the preregistration, and with the goal of achieving the same statistical power as in Study 3a, we aimed to recruit 480 university students. Aware of potential challenges in recruitment, we also preregistered that we would end data collection 2 weeks after we start, and analyze whatever number of responses we get by then, which we did. Three hundred and fifteen students from an Israeli university (36.5% male, 61.0% female; 1.9% preferred not to answer and 0.6% did not select any answer; age: $M = 27.4$, $SD = 8.8$) completed the study either for course credit or for entering a lottery. As specified in the

preregistration, we excluded from the analysis participants who failed to answer correctly at least one of the two attention checks, leading to the exclusion of 29 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.33$. We note that this study was conducted on December 2024.

Procedure

The procedure was similar to that of Study 3a with a few changes. First, the vignette (in Hebrew) was framed as happening in the past when the U.S. Army fought Al-Qaeda in Afghanistan. Second, to strengthen the moral manipulation, we referred to the Al-Qaeda members as "terrorists." Third, we used a five-item scale to measure moral judgment, as in Study 2 (Cronbach's $\alpha = .85$). Finally, we added a measure of moral motivation. We modified the measure used in Studies 1a and 2, so that it would apply also to the Al-Qaeda leader. We asked participants in both conditions to rate their agreement with the following statements about the leader they read about on a 1 (strongly disagree) to 7 (strongly agree) scale: the leader was trying to: "save as many people as possible," "maximize the number of people saved" and "reduce the number of people who die," (Cronbach's $\alpha = .93$). Finally, as attention checks, we asked participants who was the leader they read about and whether or not the leader joined the attack.

Results

Moral Judgment

A 2 (Self-Sacrifice: Yes, No) × 2 (Agent's Morality: Moral, Immoral) between-subject ANOVA revealed a main effect for Agent's morality, $F(1, 282) = 136.07$, $p < .001$, $\eta_p^2 = .325$, whereby the (moral) U.S. Army leader was judged as more moral ($M = 5.83$, $SD = 1.70$) than the (immoral) Al-Qaeda leader ($M = 3.73$, $SD = 1.38$). Replicating our previous findings, we found an effect for self-sacrifice, $F(1, 282) = 4.2$, $p = .041$, $\eta_p^2 = .015$, whereby the leaders were judged as more moral in the self-sacrifice conditions ($M = 4.94$, $SD = 2.11$) than in the no self-sacrifice conditions ($M = 4.65$, $SD = 1.59$). Importantly, this main effect for self-sacrifice was qualified by a significant Self-Sacrifice × Agent's Morality interaction,

$F(1, 282) = 6.40$, $p = .012$, $\eta_p^2 = .022$. Pairwise comparisons revealed that the U.S. Army officer was judged as more moral in the self-sacrifice condition ($M = 6.25$, $SD = 1.94$) than in the no self-sacrifice condition, ($M = 5.43$, $SD = 1.32$), $F(1, 282) = 10.64$, $p = .001$, $\eta_p^2 = .036$, whereas there was no significant difference in the judgment of the Al-Qaeda leader in the self-sacrifice ($M = 3.69$, $SD = 1.37$) and the no self-sacrifice ($M = 3.78$, $SD = 1.41$) conditions, $F(1, 282) = 0.11$, $p = .736$, $\eta_p^2 < .001$. See Figure 3.

Moral Motivation

A 2 (Self-Sacrifice: Yes, No) × 2 (Agent's Morality: Moral, Immoral) between-subject ANOVA revealed a main effect for Agent's morality, $F(1, 282) = 370.05$, $p < .001$, $\eta_p^2 = .343$, whereby the U.S. Army leader was perceived as more morally motivated ($M = 4.96$, $SD = 1.60$) than the Al-Qaeda leader ($M = 2.69$, $SD = 1.62$). Importantly, this main effect for Agent's morality was qualified by a significant Self-Sacrifice × Agent's Morality interaction, $F(1, 282) = 8.14$, $p = .005$, $\eta_p^2 = .028$. Pairwise comparisons revealed that the U.S. Army officer was perceived as more morally motivated in the self-sacrifice condition ($M = 5.37$, $SD = 1.54$) than in the no self-sacrifice condition ($M = 4.58$, $SD = 1.56$), $F(1, 282) = 8.97$, $p = .003$, $\eta_p^2 = .031$, whereas there was no significant difference in perceptions of the Al-Qaeda leader's moral motivation between the elf-sacrifice ($M = 2.55$, $SD = 1.46$) and the no self-sacrifice ($M = 2.84$, $SD = 1.78$) conditions, $F(1, 282) = 1.11$, $p = .293$, $\eta_p^2 = .004$. There was no significant main effect for self-sacrifice, $F(1, 282) = 1.825$, $p = .178$, $\eta_p^2 = .006$. See Figure 3.

Discussion

Studies 3a and 3b provide support for the causal role of motivation attribution on the effect of self-sacrifice on moral judgment. Their findings show that the effect of self-sacrifice on moral judgment is limited to cases where the perceived moral motivation is ambiguous, and thus may be affected by self-sacrifice. The results demonstrated a boundary condition for the effect of self-sacrifice: self-sacrifice did not boost moral judgment when the agent's behavior was seen as univocally immoral. This provides an important

Figure 3

Moral Judgment (Top Panel) and Perceived Moral Motivation (Bottom Panel) of Agents by Self-Sacrifice and Agent's Morality, Study 3b
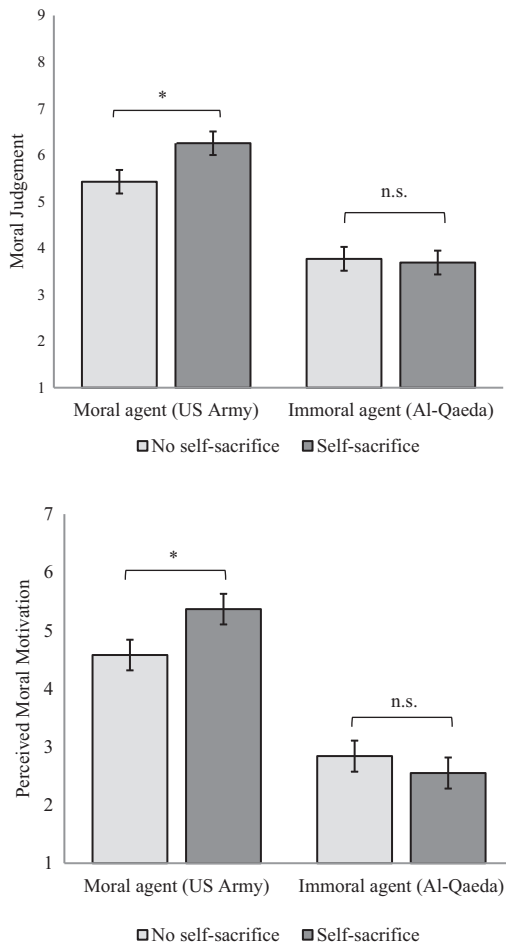
Note. Error bars denote standard errors. *n.s.* = not significant.
* $p < .05$.

qualification for our previous findings. When the immorality of the agent is clear, self-sacrifice might not affect moral judgment.

Study 4: Disentangling the Decision to Self-Sacrifice and the Sacrificial Outcome

Our previous studies revealed that self-sacrifice caused people to judge agents more positively. In those studies, the self-sacrifice manipulation involved two components: the agent's decision

to self-sacrifice (e.g., the commander's decision to join a risky mission), and the sacrificial outcome that was caused by this decision (e.g., the commander getting killed on the mission). Both these components typically coexist in self-sacrifice, but it is also possible to separate them. Past research has touched on the positive effects of mere suffering: People tend to perceive individuals who have endured suffering as more moral (Jordan & Kouchaki, 2021) and generally tend to morally elevate people who have suffered harm (Schaumberg & Mullen, 2017). But according to our theory, people's decision to risk harm to themselves is informative in and of its own for inferring moral motivation. A manipulation of the decision—while holding the outcome to the agent fixed—should affect moral judgment, via perceived moral motivation. We test this possibility in Study 4, where we independently manipulated (a) the agent's decision to put themselves in harm's way and (b) the outcome to the agent. Participants read about a U.S. Army commander who orders a risky mission and joins the soldiers or not, similar to Study 1a. We orthogonally manipulated the decision to self-sacrifice (yes or no) and the outcome (the commander died or not). Our prediction was that the commander would be judged more positively for the decision to knowingly risk his life, above and beyond the eventual outcome to him.

Method

Preregistration

https://aspredicted.org/X1V_68D.

Participants

Four hundred and eighty-one adults from the United States (50.3% male, 49.7% female; age: $M = 33.2$, $SD = 12.9$) completed the study on Prolific in exchange for 40 cents. We aimed for a sample size of 120 participants per cell. Taking into account participants who might fail attention checks and therefore be excluded from the analysis, this sample size provides a power of 0.95 to detect a medium-small effect size (Cohen's $d = 0.35$, calculated with G*Power 3.1.9.2). As specified in the preregistration, we excluded from the analysis participants who failed to answer correctly at least one of the two attention checks, leading to the exclusion of 42 participants. A sensitivity analysis using G*Power revealed that we had a

power of 0.8 to detect an effect size of Cohen's $d = 0.27$.

Procedure

Participants first answered an attention check asking them what day of the week it was and what they had for breakfast. All participants read about a Colonel in the U.S. Army who ordered his soldiers to carry out a risky attack against an Al-Qaeda terrorist cell:

> Jim Davis is a Colonel in the U.S. Army deployed in Afghanistan. A local Al-Qaeda terrorist cell has been attacking U.S. Army supply lines. If these terrorists are not dealt with, they will eventually carry out a terror attack killing some American soldiers.
>
> Colonel Davis assesses the situation and concludes that a local U.S. Army squad located nearby can take the Al-Qaeda cell out, but such an attack will be very risky and many soldiers may die.
>
> It is unclear if eliminating the terrorist group is worth the risk to the soldiers.
>
> Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack.

Participants were randomly assigned to one of four conditions, in a 2 (Self-Sacrificial Decision: Yes, No) × 2 (Outcome to Agent: Lives, Dies) between-subject design. The texts provided to participants in each of the four conditions are delineated in Table 1.

After reading the vignette, participants rated their moral judgment of the commander with the same five items as in Studies 1a and 2 (Cronbach's

$\alpha = .92$), and how moral they thought the commander's motivation was, with the same five items as in Studies 1a and 2 (Cronbach's $\alpha = .95$). Participants then answered two attention checks. The first was whether in the story they read that a U.S. colonel had died, and the second whether the colonel joined his soldiers in a mission. Finally, participants reported basic demographic information.
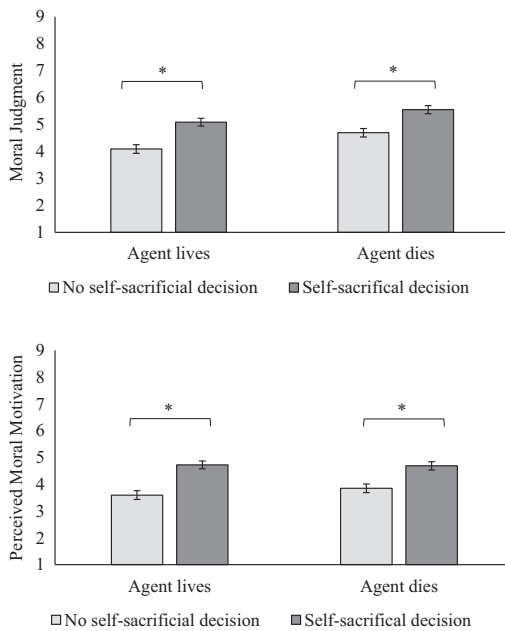
Results

Moral Judgment

A 2 (Self-Sacrificial Decision: Yes, No) × 2 (Outcome to Agent: Lives, Dies) between-subject ANOVA revealed, in line with our hypothesis, a main effect for self-sacrificial decision, $F(1, 435) = 25.91$, $p < .001$, $\eta_p^2 = .078$. The agent (Colonel Davis) was judged more positively in the conditions with self-sacrificial decision ($M = 5.32$, $SD = 1.74$) than in the conditions with no self-sacrificial decision ($M = 4.39$, $SD = 1.49$). This main effect for Self-sacrificial decision was not qualified by a Self-Sacrificial Decision × Outcome to Agent interaction, $F(1, 435) = 0.21$, $p = .651$, $\eta_p^2 < .001$, see Figure 4. The effect of self-sacrificial decision on moral judgment was thus observed with either outcome. There was also a main effect for outcome to agent, $F(1, 435) = 12.15$, $p < .001$, $\eta_p^2 = .027$, whereby the agent was judged as more moral when he died ($M = 5.12$, $SD = 1.70$) than when he lived ($M = 4.59$, $SD = 1.61$).

Table 1
Manipulations of Self-Sacrificial Decision and Outcome to Agent, Study 4

Outcome to agent	No self-sacrificial decision	Self-sacrificial decision
Agent dies	Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack. Colonel Davis does not join the squad in their mission. The squad successfully kills the terrorists. Ten of the 12 squad members are killed during the battle. While on base, Colonel Davis is hit by a stray bullet from a nearby shooting range and is killed.	Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack. Colonel Davis personally joins the squad in their mission, knowingly risking his life. The squad successfully kills the terrorists. Colonel Davis and 10 of the 12 squad members are killed during the battle.
Agent lives	Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack. Colonel Davis does not join the squad in their mission. The squad successfully kills the terrorists. Ten of the 12 squad members are killed during the battle.	Despite the risk to the soldiers, Colonel Davis orders the Army squad to carry out the attack. Colonel Davis personally joins the squad in their mission, knowingly risking his life. The squad successfully kills the terrorists. Ten of the 12 squad members are killed during the battle. Colonel Davis and two squad members survive.

Figure 4

Moral Judgment (Top Panel) and Perceived Moral Motivation (Bottom Panel) of the Agent by Self-Sacrificial Decision and Outcome, Study 4

Note. Error bars reflect standard errors.
* $p < .05$.

Moral Motivation

A 2 (Self-Sacrificial Decision: Yes, No) × 2 (Outcome to Agent: Lives, Dies) between-subject ANOVA revealed a main effect for self-sacrificial decision, $F(1, 435) = 39.75$, $p < .001$, $\eta_p^2 = .084$, whereby people perceived the agent (Colonel Davis) as more morally motivated in the conditions with self-sacrificial decision ($M = 4.71$, $SD = 1.55$) than in the conditions with no self-sacrificial decision ($M = 3.72$, $SD = 1.70$), see Figure 4. Outcome to agent had no significant effect on perception of moral motivation, $F(1, 435) = 0.49$, $p = .485$, $\eta_p^2 = .001$, and neither did Self-Sacrificial Decision × Outcome to Agent interaction, $F(1, 435) = 0.85$, $p = .358$, $\eta_p^2 = .002$. Thus, it seems that moral motivation is affected by the decision to self-sacrifice.

Mediation

To test whether perceived moral motivation mediated the link between self-sacrificial decision and moral judgment, we performed a bootstrapping mediation analysis, coding the self-sacrificial decision condition as 1 and the no self-sacrificial decision condition as −1. As predicted, the total positive effect of self-sacrificial decision on moral judgment ($b = 0.45$, $SE = 0.08$, $p < .001$) was partially mediated by an indirect effect of perceived moral motivation ($b = 0.29$, $SE = 0.05$, 95% CI [0.20, 0.38]). When accounting for the mediation by attribution of moral motivation, the direct effect of self-sacrificial decision on moral judgment was still significant ($b = 0.16$, $SE = 0.06$, 95% CI [0.04, 0.29]). This is consistent with the increase in positive moral judgment caused by self-sacrificial decision being partially driven by increased perceptions of moral motivation.

Discussion

The results of Study 4 show that a self-sacrificial decision, above and beyond outcome to the agent, causes people to judge an agent more positively, an effect plausibly mediated by an increase in perceived moral motivation. Consistent with previous work (Jordan & Kouchaki, 2021; Schaumberg & Mullen, 2017), people judged an agent who died more positively regardless of whether this was a result of a self-sacrificial decision or an accident. However, only the effect of a self-sacrificial decision was found to affect perceived moral motivation. Commander's Davis decision to join his soldiers at the risk of harming himself, and not the fact that he was killed, affected how his motivation was perceived. We also note that the effect size of the self-sacrificial decision (partial $\eta^2 = .078$) was much larger than the effect size of the outcome to agent (partial $\eta^2 = .027$), demonstrating the relative strength of the decision to self-sacrifice. In Studies 5 and 6, we expanded our investigation of the effects of self-sacrifice to a broader set of contexts and types of judgments.

Study 5: Donations

Studies 1–4 documented that people judged self-sacrificing agents more positively. Studies 5 and 6 test whether this positive judgment manifests in people's behavior toward the agents. Study 5 tested this by manipulating whether fundraisers self-sacrificed and testing if this affects how much people donate to them. Participants read either about Army reservists or about students who are fundraising for a campaign to keep

American troops in Syria (which was under debate when this study was run in May 2019). The Army reservists are self-sacrificing, as they might be called to duty and risk their lives. The students are not. We predicted that participants would donate more to the campaign when the fundraiser self-sacrificed versus not.

Method

Preregistration

https://aspredicted.org/sa8kz.pdf.

Participants

We used a larger sample size because experimental manipulations often have weaker effects on behavior intentions than on judgments (Li & Su, 2018). A power analysis revealed that 788 participants were needed participants to achieve a power of .80 to detect a small effect size for a simple correlation (Cohen's $d = 0.20$, calculated with G*Power 3.1.9.2). To account for participants who might fail the attention check, we aimed for a sample size of 800. Eight hundred and five adults from the United States completed the questionnaire on Amazon's Mechanical Turk for 20 cents (46.3% male, 53.7% female; age: $M = 38.1$, $SD = 13.2$). As specified in preregistration, we excluded from the analyses participants who failed to answer the attention check correctly, leading to the exclusion of 44 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.20$.

Procedure

Participants first answered an attention check asking them what day of the week it was and what they had for breakfast. Participants were then told that the study was about support for causes. They received a bonus of 10 cents (in addition to their basic payment for the study) and were told that they had an opportunity to donate part or all of their bonus to the cause they would read about, and that they would keep whatever part of the bonus they did not donate. Participants then read about university students who were promoting the cause of keeping U.S. Army troops in Syria. Participants were randomly assigned to one of two conditions. In the no self-sacrifice condition,

the students promoting the cause were described as studying international relations. In the self-sacrifice condition, the students promoting the cause were described as Army reservists who might personally be sent to Syria and risk their own life while deployed there. A pretest did not reveal a difference in the perceived expertise of the groups on this issue.[4] All participants read the following:

> The war on ISIS in Syria is not over. If American troops leave prematurely, the resulting instability will lead to a humanitarian crisis.
>
> Keeping the troops in Syria will certainly lead to the deaths of soldiers, but in the long run, in addition to preventing a humanitarian crisis, keeping the troops in Syria will stabilize the region, have economic benefits for the USA—and protect America's interests in the region.

In the no self-sacrifice condition, participants read:

> Although keeping the troops in Syria will put the lives of American soldiers at risk, we are well informed on the topic and strongly believe that keeping American troops in Syria is the right thing to do.

In the self-sacrifice condition, participants read the same text with the following addition:

> Although keeping the troops in Syria will put the lives of American soldiers at risk—including our own lives—we strongly believe that keeping American troops in Syria is the right thing to do.

As an attention check, participants were asked if the students promoting the cause studied international relations or were army reservists who might be sent overseas. They were then asked to what extent they thought that the students

---

[4] Participants ($N = 206$, 105 males, 100 females, 1 did not disclose; age: $M = 36.9$, $SD = 11.0$), were asked to rate which group has more expertise on whether American troops should be involved in the conflict in Syria. While 105 participants said that students of international relations (the agents in the no self-sacrifice condition) have more expertise, 101 said that Army reserve students have more expertise (the agents in the self-sacrifice condition), $\chi^2_{(1)} = 0.08$, $p = .780$, Cramer's $V = 0.02$.

We note that participants in this study might see a difference in construal level between the conditions as well. Whereas in the no self-sacrifice condition, the agents might be focusing more on the abstract, in the self-sacrifice condition, the agents might be focusing more on the concrete, low-level construal.

We also note that to strengthen our manipulation, the text participants read was generated by us, and not by army reservists or students of international relations, as participants were told.

promoting the cause were bad (1 = *extremely bad people*) or good (9 = *extremely good people*) and how much out of the 10-cent bonus they wanted to donate to this cause (0–10 cents). Participants provided demographic information and answered debriefing questions. Finally, participants were debriefed and informed that the donation question was asked just to check their response, and that the money would not be donated. They were told that instead, they would keep the full 10-cent bonus.

Results

As predicted, an independent samples *t* test revealed that participants judged the students promoting the cause as better people in the self-sacrifice condition (*M* = 6.87, *SD* = 1.71) than in the no self-sacrifice condition, *M* = 6.39, *SD* = 1.47, *t*(759) = 4.13, *p* < .001, Cohen's *d* = 0.30. Importantly, another independent samples *t* test revealed that participants donated more money (out of 10 cents) in the self-sacrifice condition (*M* = 3.07, *SD* = 3.75) than in the no self-sacrifice condition, *M* = 1.85, *SD* = 3.07, *t*(759) = 4.88, *p* < .001, Cohen's *d* = 0.36.

Finally, we tested whether the effect of self-sacrifice on donation might be mediated by moral judgment. We coded the self-sacrifice condition as 1 and the no self-sacrifice condition as −1. As noted earlier, self-sacrifice had a positive effect on donations (*b* = 0.61, *SE* = 0.12, *p* < .001). This direct effect was statistically mediated by an indirect effect of judgment (*b* = 0.18, *SE* = 0.05, 95% CI [0.09, 0.28]). When accounting for the mediation by attribution of judgment, the direct effect of self-sacrifice on donation was significant (*b* = 0.43, *SE* = 0.12, *p* < .001, 95% CI [0.20, 0.66]), consistent with a partial mediation.

Discussion

In Study 5, self-sacrifice affected participants' donation decisions. When an agent promoted a cause that was personally bad for them, participants judged them more positively and donated more money to their cause. This demonstrates that in making their donation decisions, participants cared not only about the cause itself but also about the motivation of those promoting it, providing additional support to our theory, in a new context with consequential measures.

One limitation in Study 5 is that army reservists are different than students of international relations in dimensions other than self-sacrifice. It is possible that army reservists are judged more positively than college students because they served in the past (as our results from Study 2 about past sacrifice suggest). To address this limitation, in Study 6, we used a more carefully controlled manipulation.

Study 6: Judgment of a Self-Sacrificing Politician

In Study 6, we extended our investigation and tested whether self-sacrifice affected participants' behavioral intentions toward the agents, in a controlled setting. Participants read a (fake) tweet about congressional candidate Mr. Newman, who sought to repeal the 2014 Affordable Care Act (ACA). They were told this repeal would either harm Mr. Newman by raising his health insurance premium or not affect him. Participants then rated how good Mr. Newman will be for America, their voting likelihood, and his persuasiveness. We predicted that when Mr. Newman self-sacrificed in making his policy decision, participants would rate him more favorably on all these measures. We expected self-sacrifice to affect participants' reaction to the politician, despite the fact that his potential sacrifice would be technically negligible compared to the policy's public impact. We further expected that this effect would be mediated by participants' perception of Mr. Newman as more morally motivated when he self-sacrificed.

Method

Preregistration

https://aspredicted.org/6mn9p.pdf.

Participants

We expected the same effect size as in Study 5, and therefore aimed for the same sample size of 800 participants. Seven hundred and ninety-nine adults from the United States (47.9% male, 50.4% female, 1.6% other or preferred not to disclose; age: *M* = 34.2, *SD* = 12.6) participated in the study through Prolific and were paid 30 cents as compensation. As specified in the preregistration, we excluded from the analysis participants who failed to answer at least one of the three attention

checks correctly, leading to the exclusion of 53 participants. A sensitivity analysis using G*Power revealed that we had a power of 0.8 to detect an effect size of Cohen's $d = 0.21$. Since this study may depend on current politics, we note that the data for this study were collected on June 2019.

Procedure

As an attention check, participants were asked what day of the week was yesterday and what they had for breakfast. They were told that they would read about Bill Newman who is running for Congress in their state. Participants were presented with a tweet about Mr. Newman's support of repealing the 2014 ACA. The tweet appeared real, but participants were not told whether or not it was (in fact, it was designed for the purpose of this study). Participants were randomly assigned to one of two conditions. In the self-sacrifice condition, participants read that Mr. Newman's personal premiums would increase if ACA was repealed. In the no self-sacrifice condition, they read that Mr. Newman's premiums would not change if ACA was repealed.[5] See Figure 5 for the stimuli.

Participants were asked two attention checks: (a) "Will Bill Newman pay more for health insurance if ACA is repealed?" (yes/no); (b) "Does Bill Newman want to protect ACA or to repeal it?" (Bill Newman wants to protect ACA/Bill Newman wants to repeal ACA). They were then asked to what extent they disagreed (1 = *strongly disagree*) or agreed (9 = *strongly agree*) with several statements. The first statement measured how good participants thought that Mr. Newman would be for America: "In comparison to other people from his party, if elected, Mr. Newman will be good for America." The next three statements measured the moral motivation attributed to Mr. Newman: "Mr. Newman wants to do the right thing" "Mr. Newman wants what is best for America" and "Mr. Newman has the interests of other people in mind." We created a composite moral motivation score by averaging the three items (Cronbach's $\alpha = .92$).

Two final items, preregistered as exploratory, measured how much participants were persuaded by the politician's argument: "Health insurance should be taken care of in the state level and not by the federal government," and their intention to vote for the politician "If I was in his district,

I would probably vote for Mr. Newman." Finally, participants provided demographic information.

Results

As predicted, an independent samples t test revealed that participants thought that Mr. Newman would be better for America in the self-sacrifice condition ($M = 3.52$, $SD = 2.17$) than in the no self-sacrifice condition, $M = 2.98$, $SD = 1.92$, $t(744) = 3.61$, $p < .001$, Cohen's $d = 0.26$, see Figure 6. This suggests that self-sacrifice can increase the evaluations of policy-makers, even when they hold unpopular opinions.

An independent samples t test revealed that participants reported higher voting intentions for Mr. Newman in the self-sacrifice condition ($M = 2.99$, $SD = 2.28$) than in the no self-sacrifice condition, $M = 2.38$, $SD = 1.97$, $t(744) = 3.94$, $p < .001$, Cohen's $d = 0.29$. An independent samples t test did not find a significant difference in participants' persuasion by the politician's argument in the self-sacrifice condition ($M = 3.48$, $SD = 2.44$) and the no self-sacrifice condition, $M = 3.41$, $SD = 2.38$, $t(744) = 0.44$, $p = .657$, Cohen's $d = 0.03$. Finally, another independent samples t test revealed that participants perceived Mr. Newman as more morally motivated in the self-sacrifice condition ($M = 4.11$, $SD = 2.25$) than in the no self-sacrifice condition, $M = 3.18$, $SD = 2.03$, $t(744) = 5.97$, $p < .001$, Cohen's $d = 0.43$, see Figure 6.

Next, we tested our proposed mechanism whereby the effects of self-sacrifice goodness for America and voting intentions are mediated by attribution of moral motivation. First, we tested the mediation model for goodness for America. We coded the self-sacrifice condition as 1 and the no self-sacrifice condition as -1. As noted earlier, self-sacrifice had a positive effect on goodness for America ($b = 0.27$, $SE = 0.07$, $p < .001$, 95% CI [0.12, 0.42]). This effect was mediated by an indirect effect of perceived moral motivation ($b = 0.36$, $SE = 0.06$, 95% CI [0.24, 0.48]). When accounting for the mediation by perceived of moral motivation, the direct effect of self-sacrifice on attitude toward the politician was only almost (but negatively) significant ($b =$

---

[5] To use a clean manipulation, we created the tweets participants read and invented the politician. For the manipulation to be effective, we led participants to believe they were both real.

Figure 5
Study 6 Stimuli
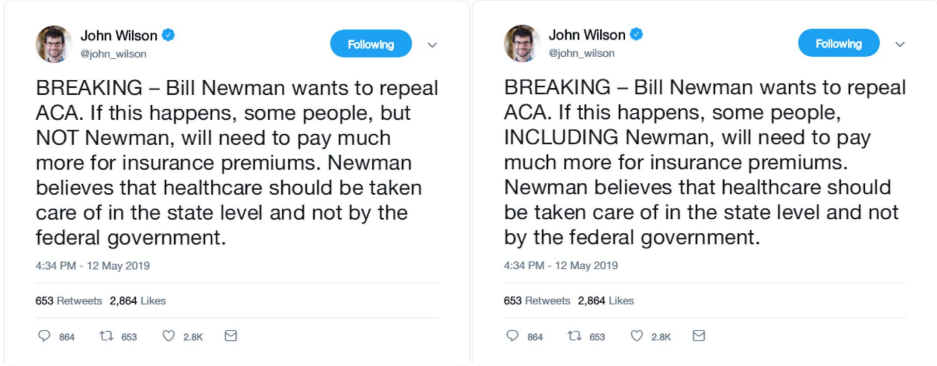
Note. Participants in Study 6 read that Mr. Newman would pay higher premiums if ACA was repealed (self-sacrifice condition, right panel) or that he would not be affected (no self-sacrifice condition, left panel). ACA = Affordable Care Act. See the online article for the color version of this figure.

−0.09, *SE* = 0.05, *p* = .065, 95% CI [−0.18, 0.01]), consistent with a partial mediation.

Second, we conducted the same mediation analysis for voting intentions. The effect of self-sacrifice on voting intentions (*b* = 0.31, *SE* = 0.08, *p* < .001, 95% CI [0.15, 0.46]) was mediated by an indirect effect of perceived moral motivation (*b* = 0.36, *SE* = 0.06, 95% CI [0.24, 0.49]). When accounting for the mediation by perceived moral motivation, the direct effect of self-sacrifice on voting intentions was not significant (*b* = −0.06, *SE* = 0.05, *p* = .270, 95% CI [−0.15, 0.04]).

Figure 6
Results of Study 6
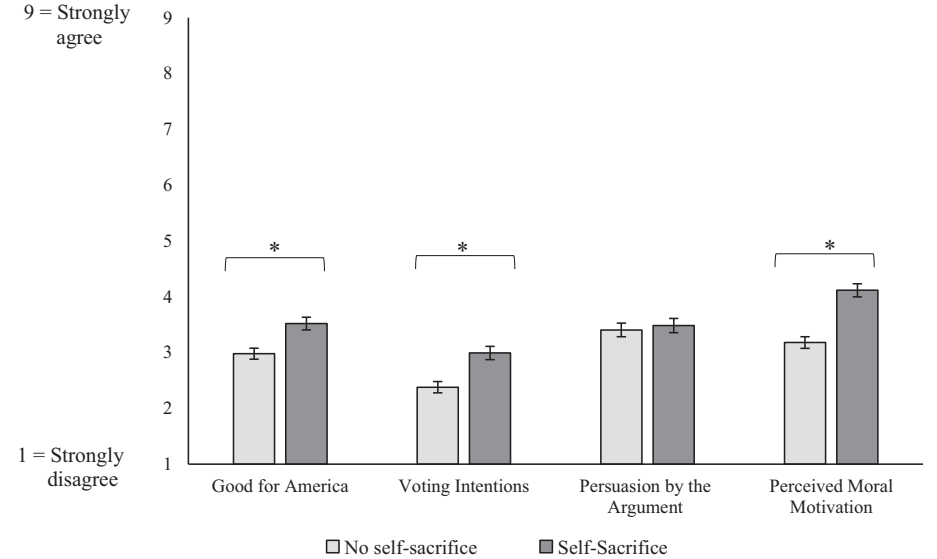
Note. Participants' judgments of the agent (Mr. Newman) are presented by condition. Error bars reflect standard errors.
* *p* < .05.

Discussion

The results from Study 6 demonstrate the impact of self-sacrifice on attitudes toward a politician supporting a relatively unpopular policy proposal. Participants thought the politician would be better for America and even reported that they would be more likely to vote for him (but not that they were more persuaded by the politician's arguments), when the politician made a policy decision that was against his personal interest. We also found that perceived moral motivation mediated these effects, plausibly such that when a politician supported a contested policy that harmed him, participants were more likely to think that he did so because he sincerely believed that it would have good outcomes.

Internal Meta Analyses

To summarize the effect sizes, we found across studies and evaluate the overall effect size (Goh et al., 2016), we conducted two mini meta-analyses. The first mini meta-analysis delineates the effect of self-sacrifice on moral judgment. We used an restricted maximum likelihood random-effect model and included all studies that measured moral judgment. For Studies 3a and 3b, we excluded the two boundary conditions where the motivation of the agent was clearly immoral and we did not predict a self-sacrifice effect. This analysis revealed an estimated effect size of Cohen's $d = 0.567$, $Z = 6.709$, $p < .001$. See Figure 7.

The second mini meta-analysis summarized the effect of self-sacrifice on perceived moral motivation. We used an restricted maximum likelihood random-effect model and included all studies that measured perceived moral motivation. For Study 3b, we excluded the two boundary conditions where the motivation of the agent was clearly immoral, and we did not predict an effect for self-sacrifice. This analysis revealed an estimated effect size of Cohen's $d = 0.52$, $Z = 5.791$, $p < .001$. See Figure 7.

General Discussion

Although we know from previous research that moral motivation shapes moral judgment (Carlson et al., 2022), it is unclear which cues people use to infer about moral motivation when the action is morally ambiguous. Here, we demonstrate that

self-sacrifice is such a cue. When agents decide to take actions for which they know they might pay a price, people infer that their motivation is moral and accordingly judge them more positively. Six studies provided consistent evidence for this theoretical model (see Figure 7 for two internal meta-analyses).
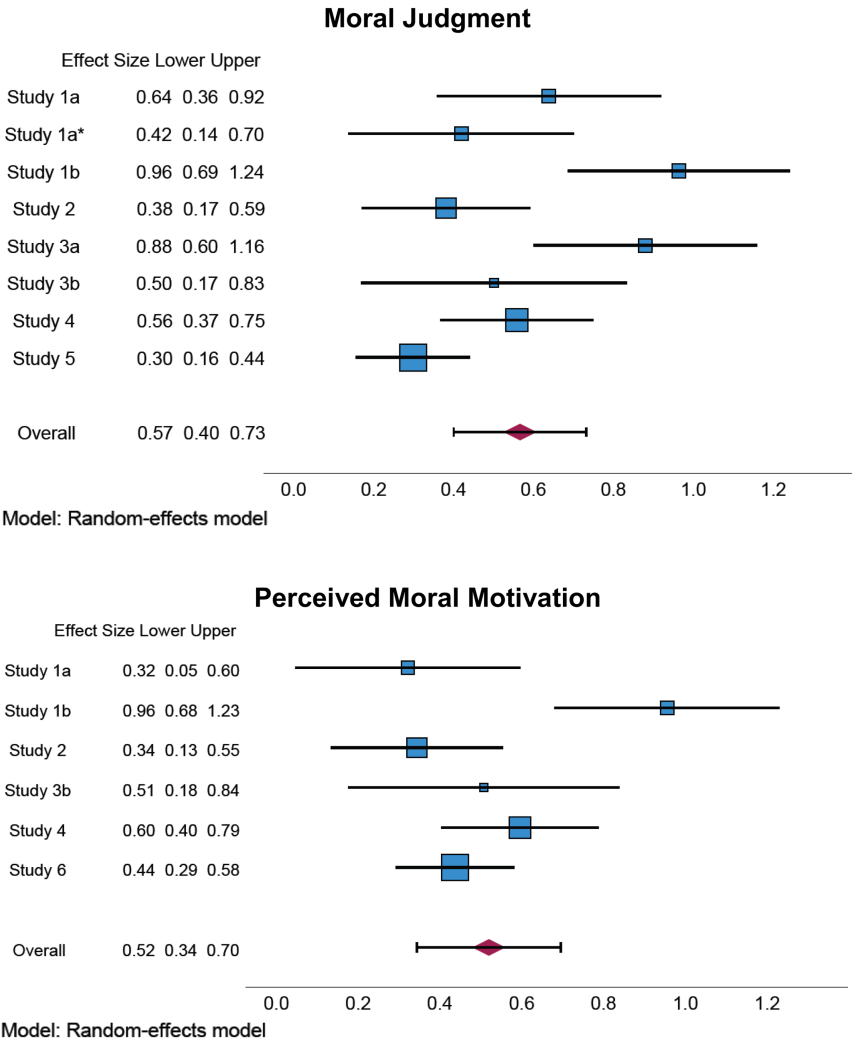
In line with our theoretical model, the more positive moral judgment of self-sacrificing agents was accounted for by a perception of them as having more moral motivations (Studies 1 and 2). Furthermore, a given behavior of an agent was seen as more moral if they had self-sacrificed in a different occasion in the past, because this reflected on their motivation (Study 2). The self-sacrifice effect was not obtained when the agent's motivation was clearly immoral (Studies 3a and 3b), providing important support for the causal role of perceived motivation and highlighting a critical boundary condition. The effect of the decision to self-sacrifice is also independent of the eventual outcome. An agent who intentionally puts herself in harm's way is perceived as having a more moral motivation and therefore judged as more moral, even if she did not end up being harmed (Study 4). Self-sacrifice further affected related behavior. When the supporters of a controversial cause were said to self-sacrifice, participants donated more money to the cause they promoted (Study 5). When a politician was said to make a personal sacrifice, participants evaluated him more positively and reported greater intentions to vote for him (Study 6). Taken together, our findings demonstrate the importance of perceived moral motivation in moral judgment and highlight how self-sacrifice affects inferences about moral judgment.

Theoretical and Practical Implications

Decades of research highlight the role of motivation attribution in person perception and moral judgment (e.g., Heider, 1958; Reeder et al., 2002). When actions have clear positive or negative outcomes, moral judgments are straightforward, with positive outcomes linked to moral motives and negative ones to immoral motives (Carlson et al., 2022). This article extends prior work by examining cases where actions have ambiguous outcomes, finding that self-sacrifice serves as a key cue for moral attribution.

Our findings complement research on tainted altruism, where agents who help others are judged

Figure 7

Internal Meta-Analyses on the Effect of Self-Sacrifice on Moral Judgment (Top Panel) and Perceived Moral Motivation (Bottom Panel)

Moral Judgment

	Effect Size	Lower	Upper
Study 1a	0.64	0.36	0.92
Study 1a*	0.42	0.14	0.70
Study 1b	0.96	0.69	1.24
Study 2	0.38	0.17	0.59
Study 3a	0.88	0.60	1.16
Study 3b	0.50	0.17	0.83
Study 4	0.56	0.37	0.75
Study 5	0.30	0.16	0.44
Overall	0.57	0.40	0.73

Model: Random-effects model

Perceived Moral Motivation

	Effect Size	Lower	Upper
Study 1a	0.32	0.05	0.60
Study 1b	0.96	0.68	1.23
Study 2	0.34	0.13	0.55
Study 3b	0.51	0.18	0.84
Study 4	0.60	0.40	0.79
Study 6	0.44	0.29	0.58
Overall	0.52	0.34	0.70

Model: Random-effects model

Note. Effect sizes for Studies 3a and 3b are presented only for the moral agent conditions. Effect sizes for Study 4 are across outcome conditions. Study 1a: Replication study is labeled Study 1a* in the figure. Box size reflects the sample size of the study. Lines depict 0.95 confidence intervals. See the online article for the color version of this figure.

negatively if they also benefit themselves (Alempaki et al., 2023; Newman & Cain, 2014). Note that an agent being judged more negatively for self-benefitting is not necessarily a symmetrical phenomenon to an agent being judged more positively for self-sacrificing. Indeed, moral judgments often display asymmetries between positive and negative behaviors (Klein & Epley, 2014; Knobe, 2003). Our research fills this gap by showing that self-sacrifice influences moral judgment in an opposite direction to self-benefit: People are judged positively when self-sacrificing. Future work could systematically compare the effects of self-sacrifice and self-benefit on moral evaluation and their magnitude.

Our work also contributes to the broader discussion of sacrificial moral acts, where agents typically sacrifice *others* for the greater good.

Past work revealed that participants interpret "utilitarian" decisions in moral dilemmas (e.g., harming one person to save many) either as acts of cold-hearted callousness or as difficult humanitarian decisions (Kahane et al., 2018). Our findings suggest that whereas sacrificing others does not reveal much on the agent's morality, self-sacrifice is far more informative. An agent who pushed a large person off a bridge to prevent a runaway trolley from killing five workmen down the tracks might be a real utilitarian motivated to impartially save as many people as possible or an untrustworthy psychopath (Bartels & Pizarro, 2011; Everett et al., 2016; Kahane et al., 2015, 2018). However, if the agent jumped from the bridge to block the runway trolley with their own body, the motivational ambiguity disappears. Our work further suggests that in some cases, self-sacrifice will cause a more positive moral judgment even when by self-sacrificing the overall utility of the action is lower (e.g., more people died).

Our research expands on previous work on self-sacrifice. Some of the previous work investigates the psychological traits that increase the likelihood of self-sacrificing (Bélanger et al., 2014). Other work found that people are more willing to donate to a charity or engage in other prosocial behavior (Sachdeva et al., 2015), when it requires them to personally self-sacrifice. For example, donors exhibited greater willingness to contribute to charity when the fundraiser required them to run—an activity that is more effortful and physically demanding—compared to when it involved participating in a picnic (Olivola & Shafir, 2013). Other research examining judgments of one-time donations versus repeated donations found that repeated donations, which were associated with self-sacrifice, were perceived as more moral than one-time donations of the same total amount (Basu, 2021). Our work advances this research by experimentally demonstrating the causal role of self-sacrifice in shaping how people judge others and by highlighting the influence of perceived moral motivation in this process.

In this article, we focused on how observers judge agents' actions. Our findings have implications for the agents themselves. Generally, people's beliefs about how others judge them shape their behavior (Guzikevits & Choshen-Hillel, 2022; Leary & Kowalski, 1990). In the current context, our findings imply that agents may expect observers to judge them more favorably when they self-sacrifice. Agents may thus deliberately choose to self-sacrifice to signal that their motivation is good and to reduce the risk of being morally condemned. Agents may be more willing to self-sacrifice when they are more concerned about appearing moral, and when optics are more important for them. Some prior research is consistent with this hypothesis. People are willing to place themselves at a disadvantage to signal their fairness (Choshen-Hillel et al., 2015; Shaw et al., 2016) and are willing to incur some monetary costs to signal their honesty (Choshen-Hillel et al., 2020; Guzikevits & Choshen-Hillel, 2022). People may even sacrifice the interests of their close friends when being scrutinized to signal that they are being impartial (Shaw et al., 2018). Therefore, people might strategically engage in unnecessary self-sacrifice to maintain their moral reputation. On the flip side, agents may be less likely to self-sacrifice when they expect that observers will not know about it.

The concept of self-sacrifice in this article aligns with the theory of (costly) signaling. According to signaling theory, costly behaviors can serve as credible indicators of underlying traits or intentions—such as sincerity and commitment (Henrich, 2009)—because they are difficult to fake (Barker et al., 2019; Bliege Bird & Smith, 2005; Zahavi & Zahavi, 1997). Our work on self-sacrifice demonstrates how the principle of costly signaling might manifest itself in the moral domain, by serving as a cue to moral motivation.

Finally, our findings have practical implications. People are, perhaps justifiably, suspicious of other's motivations (Ratner & Miller, 2001). Self-sacrifice may be used as an effective tool to dispel such suspicions. For example, politicians, managers, and organizations can use self-sacrifice to elicit support and protect their reputation when enacting unpopular decisions. The public may show greater support for a wealth tax when it is promoted by wealthy people who would need to pay higher taxes themselves, and employees may view their chief executive officer's decision more positively if the chief executive officer is negatively affected by a company policy they passed.

Limitations and Future Directions

Participants in our studies were mostly from North America, although we did replicate the self-sacrifice effect in samples of Israeli college students (Study 1a: replication and Study 3b). Still,

the cultural generalizability of our findings needs to be systematically tested. It is possible that in religions and cultures where mental-state perceptions are less important for moral judgments (Cohen & Rozin, 2001; Curtin et al., 2020), self-sacrifice might affect moral judgment less, and perhaps even not at all.

Some of our studies, that used the scenario of a commander self-sacrificing by joining their troops on a risky mission (Studies 1a, 2–4), have another limitation. It is possible that participants had a more positive moral judgment of the agent, because they perceived him as courageous and as performing his duty. However, we note that our studies include contexts where courage and duty are unlikely to explain the effect, such as a student opposing affirmative action (Study 1b) and a politician who wants to repeal the ACA (Study 6).

Another possible boundary condition is the ratio between the level of self-sacrifice and the overall utility of the act. When making moral decisions, people are sensitive to the harm-benefit ratio (Trémolière & Bonnefon, 2014). Furthermore, people might not judge a self-sacrificing agent more positively if the harm–benefit ratio is sufficiently large, whether because many people are harmed or because the benefits are too small. For example, a person who sacrifices their life to save another person from a minor injury might not be considered moral, but rather weird. Future research is needed to systematically study how the ratio between the degree of self-sacrifice and the overall utility of the action shapes moral judgments.

In some of our studies, the no self-sacrifice condition included some benefit to the agent. For example, in Study 1b, where we manipulated whether a male of a female student oppose gender-equality affirmative action, participants might see the male as benefitting from his support of the policy. In these cases, the observed effect could have been driven by the benefit in the no self-sacrifice condition, rather than by the cost in the self-sacrifice condition. We note that in other studies such as Study 6, there was no benefit associated with the no self-sacrifice condition (participants read about a politician wanting to repeal the ACA, which would not benefit him directly), supporting our conclusion on the role of the cost to self.

Conclusion

When people form judgments of others, they care not only about the actions of others and the consequences of those actions. They also care about *why* they behaved the way they did—what motivated them. Since other people's minds are ultimately inaccessible, people must use observable behavior cues in the quest to interpret other people's actions. Our present findings show that in cases where actions are morally ambiguous, one powerful cue used to infer the motivation of others is self-sacrifice.

References

Alempaki, D., Isoni, A., & Read, D. (2023). Tainted nudge. *Organizational Behavior and Human Decision Processes*, *176*, Article 104244. https://doi.org/10.1016/j.obhdp.2023.104244

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574. https://doi.org/10.1037/0033-2909.126.4.556

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755–1762. https://doi.org/10.1177/0956797613480507

Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 396–409). Wiley. https://doi.org/10.1002/9780470561119.socpsy001008

Barker, J. L., Power, E. A., Heap, S., Puurtinen, M., & Sosis, R. (2019). Content, cost, and context: A framework for understanding human signaling systems. *Evolutionary Anthropology*, *28*(2), 86–99. https://doi.org/10.1002/evan.21768

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Basu, S. (2021). Framing an altruistic action in periodic (versus aggregate) terms reduces people's moral evaluation of the act and the actor. *Journal of Experimental Social Psychology*, *97*, Article 104205. https://doi.org/10.1016/j.jesp.2021.104205

Bélanger, J. J., Caouette, J., Sharvit, K., & Dugas, M. (2014). The psychology of martyrdom: Making the ultimate sacrifice in the name of a cause. *Journal of Personality and Social Psychology*, *107*(3), 494–515. https://doi.org/10.1037/a0036855

Bigman, Y. E., Gray, K., & Choshen-Hillel, S. (2025). Materials and data for: Morally motivated? People use self-sacrifice as a cue for moral character. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/5RDXU

Bigman, Y. E., & Tamir, M. (2016). The road to heaven is paved with effort: Perceived effort amplifies moral judgment. *Journal of Experimental Psychology:*

General, *145*(12), 1654–1669. https://doi.org/10 .1037/xge0000230

Bliege Bird, R., & Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology*, *46*(2), 221–248. https:// doi.org/10.1086/427115

Carlson, R. W., Bigman, Y. E., Gray, K., Ferguson, M. J., & Crockett, M. J. (2022). How inferred motives shape moral judgment. *Nature Reviews Psychology*, *1*(8), 468–478. https://doi.org/10.1038/s44159-022-00071-x

Carlson, R. W., & Zaki, J. (2018). Good deeds gone bad: Lay theories of altruism and selfishness. *Journal of Experimental Social Psychology*, *75*, 36–40. https://doi.org/10.1016/j.jesp.2017.11.005

Chakroff, A., Russell, P. S., Piazza, J., & Young, L. (2017). From impure to harmful: Asymmetric expectations about immoral agents. *Journal of Experimental Social Psychology*, *69*, 201–209. https://doi.org/10.1016/j.jesp.2016.08.001

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology*, *109*(2), 210–231. https://doi.org/10.1037/pspa0000028

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020). Lying to appear honest. *Journal of Experimental Psychology: General*, *149*(9), 1719–1735. https:// doi.org/10.1037/xge0000737

Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, *81*(4), 697–710. https://doi.org/ 10.1037/0022-3514.81.4.697

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological & Personality Science*, *4*(3), 308–315. https://doi.org/10.1177/194855061245 7688

Curtin, C. M., Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., Kanovsky, M., Kushnick, G., Laurence, S., Pisor, A., Scelza, B., Stich, S., von Rueden, C., & Henrich, J. (2020). Kinship intensity and the use of mental states in moral judgment across societies. *Evolution and Human Behavior*, *41*(5), 415–429. https://doi.org/10.1016/j.evolhu mbehav.2020.07.002

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https:// doi.org/10.1016/j.cognition.2008.03.006

Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*(6), 772–787. https://doi.org/10.1037/ xge0000165

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. https://doi.org/ 10.1037/a0034726

Guzikevits, M., & Choshen-Hillel, S. (2022). The optics of lying: How pursuing an honest social image shapes dishonest behavior. *Current Opinion in Psychology*, *46*, Article 101384. https://doi.org/ 10.1016/j.copsyc.2022.101384

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley. https://doi.org/10.1037/10628-000

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, *30*(4), 244–260. https://doi.org/10.1016/j.evolhumbehav.2009.03.005

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin*, *38*(1), 52–62. https://doi.org/10.1177/0146167211430232

Jordan, J. J., & Kouchaki, M. (2021). Virtuous victims. *Science Advances*, *7*(42), Article eabg5902. https://doi.org/10.1126/sciadv.abg5902

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193–209. https://doi.org/10.1016/j.cognition.2014.10.005

Klein, N., & Epley, N. (2014). The topography of generosity: Asymmetric evaluations of prosocial actions. *Journal of Experimental Psychology: General*, *143*(6), 2366–2379. https://doi.org/10 .1037/xge0000025

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190–194. https://doi.org/10.1093/analys/63.3.190

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34–47. https:// doi.org/10.1037/0033-2909.107.1.34

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107–117. https://doi.org/10.1016/j.jesp.2014.03 .005

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88–106. https://doi.org/10.1016/j.obhdp.2014.10.007

Li, N., & Su, L. Y.-F. (2018). Message framing and climate change communication: A meta-analytical review. *Journal of Applied Communications*, *102*(3). https://doi.org/10.4148/1051-0834.2189

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2022). Moral judgment as categorization (MJAC). *Perspectives on Psychological Science*, *17*(1), 131–152. https://doi.org/10.1177/1745691621990636

Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, *54*(12), 1053–1060. https://doi.org/10.1037/0003-066X.54.12.1053

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, *74*(1), 53–62. https://doi.org/10.1037/0022-3514.74.1.53

Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science*, *25*(3), 648–655. https://doi.org/10.1177/0956797613504785

Olivola, C. Y., & Shafir, E. (2013). The martyrdom effect: When pain and effort increase prosocial contributions. *Journal of Behavioral Decision Making*, *26*(1), 91–105. https://doi.org/10.1002/bdm.767

Pizarro, D. A., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, *14*(3), 267–272. https://doi.org/10.1111/1467-9280.03433

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). American Psychological Association. https://doi.org/10.1037/13091-005

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Ratner, R. K., & Miller, D. T. (2001). The norm of self-interest and its effects on social action. *Journal of Personality and Social Psychology*, *81*(1), 5–16. https://doi.org/10.1037/0022-3514.81.1.5

Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, *83*(4), 789–803. https://doi.org/10.1037/0022-3514.83.4.789

Rohrer, J. M. (2024). Causal inference for psychologists who think that causal inference is not for them. *Social and Personality Psychology Compass*, *18*(3), Article e12948. https://doi.org/10.1111/spc3.12948

Sachdeva, S., Iliev, R., Ekhtiari, H., & Dehghani, M. (2015). The role of self-sacrifice in moral dilemmas. *PLOS ONE*, *10*(6), Article e0127409. https://doi.org/10.1371/journal.pone.0127409

Schaumberg, R. L., & Mullen, E. (2017). From incidental harms to moral elevation: The positive effect of experiencing unintentional, uncontrollable, and unavoidable harms on perceived moral character. *Journal of Experimental Social Psychology*, *73*, 86–96. https://doi.org/10.1016/j.jesp.2017.06.016

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2016). The development of inequity aversion: Understanding when (and why) people give others the bigger piece of the pie. *Psychological Science*, *27*(10), 1352–1359. https://doi.org/10.1177/0956797616660548

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2018). Being biased against friends to appear unbiased. *Journal of Experimental Social Psychology*, *78*, 104–115. https://doi.org/10.1016/j.jesp.2018.05.009

Trémolière, B., & Bonnefon, J.-F. (2014). Efficient kill-save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*(7), 923–930. https://doi.org/10.1177/0146167214530436

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81. https://doi.org/10.1177/1745691614556679

Uhlmann, E. L., Zhu, L. L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology*, *44*(1), 23–29. https://doi.org/10.1002/ejsp.1987

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. Oxford University Press. https://doi.org/10.1093/oso/9780195100358.001.0001