Recognizing and Correcting Positive Bias: The Salient Victim Effect

Emily M. Zitek [1]

Laura M. Giurge [2]

Isaac H. Smith [3]

[1] School of Industrial and Labor Relations, Cornell University, United States

[2]London School of Economics and Political Science, United Kingdom

[3] Marriott School of Business, Brigham Young University, United States

Article accepted for publication at *Journal of Experimental Social Psychology.*

This article may not exactly replicate the final version published in the APA journal.

Corresponding Author: Emily Zitek, School of Industrial and Labor Relations, Cornell University, Ives Faculty Building, Ithaca, NY, 14853, United States. Email: emily.zitek@cornell.edu

Abstract

People seem to have stronger disapproving reactions when they have unfairly suffered from bias than when they have unfairly benefited from it (i.e., they seem less concerned when they have experienced *positive bias*). Is this because people do not care about the consequences of bias if it has positively affected them, or is it because they fail to notice positive bias? We argue that it is the latter, and that increasing awareness of a victim who has been harmed can "remove the blinders" of the beneficiary of bias. Across seven pre-registered studies of American participants, we tested the effect of a salient victim on people who have experienced positive bias. Our results show that when a victim has been made salient, beneficiaries of bias are more likely to recognize and condemn the positive bias, and they are also more likely to act to correct it. We found this *salient victim effect* when people reflected on their own positive treatment in society, when they benefited from favoritism in interpersonal interactions, and when they imagined benefiting from nepotism. The effect emerged with both direct and indirect manipulations of the victim. Moreover, the presence of a salient victim spurred more action in those who experienced positive bias even when there was a personal cost. We discuss the contributions of our research to the fairness, morality, and bias literatures.

Keywords: bias, unfairness, victim, positive outcome, favoritism

Recognizing and Correcting Positive Bias: The Salient Victim Effect

Biases related to age, race, socio-economic status, gender, familiarity, and even physical appearance can result in the unjustified mistreatment of others. However, while we often focus on the victims of such bias, there are also many beneficiaries of bias. Consider unqualified employees who receive a competitive promotion because their boss is a close friend, or drivers who get a "warning" instead of a speeding ticket because they have the "right look." These people have experienced what we label *positive bias* (i.e., receiving an unmerited favorable outcome due to preferential treatment), and although most people endorse fairness as a foundation of morality (Graham et al., 2011), recipients of positive bias seem unlikely to push back against it. For example, students often complain that they were graded too harshly because their teacher is "playing favorites," but they rarely (if ever) complain that they were graded too leniently for the same reason. Is this seeming moral hypocrisy due to people's lack of caring that others are harmed by biases if they themselves have benefited, or is it that they fail to recognize—or perhaps are willfully ignorant—that they are benefiting from an unfair positive bias in the first place? In this paper, we provide evidence for the latter, arguing that positive bias can remain hidden in people's ethical blind spots (see Sezer et al., 2015). Drawing on the Theory of Dyadic Morality (TDM; Schein & Gray, 2018), we argue that one way to help people recognize—and then act to correct—positive bias is by making the victim(s) of positive bias salient.

Reactions to Bias

People subjected to unfair, negative events such as underpayment, disrespect, or biased treatment often have intense reactions. They not only experience negative emotions such as anger, but they may also exhibit a stress response (Dion & Earn, 1975; Jamieson et al., 2013;

Mikula et al., 1998; Miller, 2001; Richman & Leary, 2009; Tabibnia et al., 2008). Victims of biased treatment and other injustices might try to retaliate, compensate for the negative outcome, or otherwise take action to correct it (e.g., Barclay et al., 2005; Walster et al., 1978; Zitek et al., 2010). For example, victims of bias in the workplace might confront the biased individual if they think this person can change, which can lead to positive outcomes (Rattan & Dweck, 2018). In short, there is clear evidence that victims of biases (and other injustices) are bothered by what has happened to them and often seek to somehow address the perceived injustices.

However, we do not expect beneficiaries of bias to respond in the same way as victims of bias. People process and react differently to positive and negative events (e.g., Baron & Hershey, 1988; Baumeister et al., 2001; Davidai & Gilovich, 2016; Rozin & Royzman, 2001; Taylor, 1991), including events that are inequitable (e.g., Mowday, 1996; Shore, 2004). For example, people do not seem to have as strong of a desire to correct inequity when they have received too much rather than too little (e.g., Austin & Walster, 1975; Gray et al., 2014). An undeserved outcome feels different if it is positive for the recipient rather than negative (Feather & McKee, 2009). To directly test how people react to positive versus negative bias, we conducted some pilot studies (see the Supplemental Online Materials, SOM).

In Pilot Study 1, we asked some participants to imagine that they had to pay *less* for a hotel room because the hotel manager didn't think they "looked like" they would do anything destructive (i.e., positive bias). These participants rated the hotel manager's behavior as fair and ethical. However, when other participants were asked to imagine that they had to pay *more* because they "looked like" they would do something destructive, they considered this to be quite unfair and unethical, even though the same method (predicting someone's behavior based on their appearance) was used in both cases to make the decision about the cost of the room (see

also Pilot Study 2). In line with these results, other researchers have also argued that positive bias is often invisible to the beneficiaries (Phillips & Lowery, 2018). Because positive bias confers welcome benefits, and because people want to believe they are moral individuals who have not contributed to any unfairness (Dunning, 2007; Mazar et al., 2008), a beneficiary of bias is likely motivated to ignore the bias (see Gino et al., 2010). From this perspective, it isn't that beneficiaries of bias do not care about the victims; they just may not realize that anything immoral has happened or that anyone has suffered unfairly. To better understand why positive bias might be hard to recognize, we turn to TDM (Schein & Gray, 2018).

According to TDM, an act will be morally condemned if it violates a norm, triggers negative affect, and results in dyadic harm where an intentional agent causes harm to a vulnerable victim (see also Gray et al., 2022; Nichols, 2002; Schein et al., 2016). Positive bias might be hard to recognize (and condemn) because beneficiaries of bias do not identify a norm violation, feel any negative affect, or notice any harm.

First, recipients of positive bias will likely generate an explanation for why their positive outcome was fair, which may cause them to fail to identify a norm violation. Indeed, people often find a way to justify their good outcomes, even when undeserved (Ellard & Bates, 1990; Gaucher et al., 2010). They do so by making internal attributions for these outcomes (i.e., the self-serving bias; Mezulis et al., 2004) or otherwise thinking about them in ways that help maintain their belief in a just world (Hafer & Bègue, 2005). For example, upper class individuals might bring up their hardships or hard work to make their privilege seem merited (Phillips & Lowery, 2020).

Second, positive outcomes elicit positive affect, which leads to less rumination and less desire for corrective action than negative affect does (e.g., Gray et al., 2014; Moberly &

Watkins, 2008). Negative emotions highlight the potential immorality of the situation (Avramova & Inbar, 2013), and without these negative emotions, the bias might be missed. Indeed, people filter their experiences through their emotions: When people are in a positive mood, they are more likely to see the world through "rose-colored glasses" (Clore, 1994; Van Kleef, 2009) and thus less likely to question the fairness of the favorable outcome that put them in a positive mood.

Third, egocentric biases (e.g., Gilovich et al., 1998) may prevent beneficiaries of bias from noticing any harm done to others because their own outcome was positive. Indeed, in Pilot Study 1 (see SOM), participants who imagined experiencing positive bias from the hotel manager failed to recognize the potential harm to others due to the hotel manager's behavior. People do not think as carefully about an outcome they like as compared to one they do not like (e.g., Ditto & Lopez, 1992; Ditto et al., 2009), and, as a result, they might not realize that someone else was harmed when they receive an unmerited favorable outcome. Without any noticeable harm, people will be less likely to view the action as unethical (Gino et al., 2008).

Recognizing Positive Bias

Given that people seem to be less likely to recognize and condemn positive bias, an important question emerges: When *will* people recognize it, given that they are motivated (often unconsciously) to keep positive bias in their ethical blind spots? When someone benefits from bias, there is often a victim lurking in the background (e.g., when someone receives a promotion from a family member, there are others who miss out on that promotion), but people may need to have their attention drawn toward this victim to recognize that they benefited from positive bias. We argue that a salient victim makes it more likely that people will identify a norm violation, feel negative affect, and notice the harm of positive bias, and thus, in line with TDM (Schein &

Gray, 2018), become aware that something potentially immoral has happened (i.e., they will recognize the unfairness due to positive bias).

First, a salient victim likely demonstrates that there was a norm violation whereby some people received unmerited advantages over others. Consistent treatment across individuals is a key component of fairness judgments (e.g., Adams, 1965; Cropanzano et al., 2007; Matta et al., 2017), and when victims of bias become salient, people might be more likely to realize that not everyone was treated in the same way.

Second, a salient victim increases the likelihood that a beneficiary of bias will experience negative affect (e.g., Erlandsson et al., 2015; Kogut & Ritov, 2005). Whereas people usually feel good when they receive a positive outcome, they often also feel bad for others who receive a negative outcome. People might feel sadness, distress, anger, sympathy, or empathic concern in response to the victim's poor outcome (e.g., Zitek & Jordan, 2021), or guilt because of their benefits relative to the victim's (Brockner et al., 1986; Mattila et al., 2013).

Third, a victim focuses the beneficiary of bias on the harm of the situation. Research on egocentrism shows that people often think about things from their own perspective and fail to adjust for the perspective of others (e.g., Bohns, 2016; Gilovich et al., 1998; Gilovich et al., 2000; Giurge & Bohns, 2021). We assert that highlighting a salient victim can help individuals overcome their egocentric biases and take the perspective of others when it comes to positive bias. This, in turn, may make people more likely to perceive the original action (e.g., the family member's promotion over unrelated yet more qualified individuals) as biased and unfair. Indeed, research confirms that recognizing harm done to others can prevent people from morally disengaging in situations of self-interest (Kish-Gephart et al., 2014).

In short, people often engage in motivated blindness, whereby they disregard the unethical behavior of others that they are benefiting from (Gino et al., 2010; Sezer et al., 2015). A victim may prevent this motivated blindness by making it difficult for people to overlook the norm violation, not feel bad about what happened, or miss the harm. Integrating the above research and reasoning, we hypothesize that people will be more likely to recognize positive bias when the victims who have been harmed by their good fortune are made salient.

Taking Corrective Action

Of course, to counteract positive bias, people need to do more than just recognize it—they also need to speak up about it or take other action to correct it. Beneficiaries of bias may be reluctant to take corrective action, as they may not want to irritate or embarrass someone who helped them or lose out on the positive treatment they received (see Carlson et al., 2022). However, failing to counteract positive bias may constitute moral hypocrisy (Batson & Thompson, 2001)—acting in a manner inconsistent with espoused moral values. Sometimes moral hypocrisy may be unintentional, a form of self-deception (Batson et al. 1999), wherein people fail to recognize that their behavior violates their moral values. At other times, moral hypocrisy can be intentional, as people perceive that the benefits of violating their values outweigh the costs, leading them to act in ways contrary to their values (Batson & Thompson, 2001). We expect that a salient victim will increase a person's awareness of positive bias (reducing unintentional moral hypocrisy) and, subsequently, amplify the costs of not counteracting it (reducing intentional moral hypocrisy).

Once the salient victim gets people to recognize that they have unfairly benefited from bias, they may be more likely to take corrective action because of their desire to maintain a positive moral self-view (Aquino & Reed, 2002; Dunning, 2007; Mazar et al., 2008). Fairness is

a widespread and nearly universal moral value (Graham et al., 2011; Peterson & Seligman, 2004). Not speaking up about (and even benefiting from) unfairness can cause ethical dissonance, a psychological discomfort arising from an inconsistency between one's behavior and one's moral values (Barkan et al., 2012). A salient victim may make it hard for people to avoid comparing their behavior to a relevant moral standard (Batson et al., 1999). Thus, we argue that a salient victim makes beneficiaries of bias recognize the moral problems inherent in the situation, and as a result, they will be more likely to take action to solve the problems (see also Jones, 1991)—and therefore less likely to engage in unintentional moral hypocrisy.

Moreover, after the salient victim has made the positive bias readily apparent, the costs of ignoring the bias increase (Batson & Thompson, 2001), making the beneficiaries feel like they must counteract the bias to avoid being or appearing immoral or hypocritical (Batson et al., 1999; Batson & Thompson, 2001; Lönnqvist et al., 2014; Monin & Merritt, 2012). The victim may compel people to act because it makes it harder for them to make self-serving justifications for why inaction is okay (Shalvi et al., 2015), or because they believe that others will now expect them to do the right thing. The negative personal and public consequences of not taking corrective action thus increase, making intentional moral hypocrisy more costly.

Consistent with the above points on the relationship between recognizing and correcting unfairness, research has shown that people are more likely to support action to reduce economic inequality when they perceive the inequality as an injustice (Dietze & Craig, 2021). Relatedly, when people are made aware of their own biases, they try to correct the problem by acting more unbiased (Perry et al., 2015), and we expect that they might also try to correct the problem when they notice bias in others. Research has documented that people are sometimes willing to take a stand against immoral behavior even at a cost to themselves (Cao et al., 2019; Fehr & Gatchter,

2002), and this might happen for positive bias when there is a salient victim. Specifically, we hypothesize that the presence of a salient victim will lead individuals who benefit from positive bias to be more likely to take corrective action because they will have a higher likelihood of recognizing the bias in the first place, which increases the cost of not addressing it.

Overview of Studies

In seven pre-registered studies, we examined our proposed *salient victim effect* on people's likelihood of recognizing and acting against positive bias across a variety of situations. In Study 1, we examined whether White Americans exposed to salient victims of racial profiling are more likely to report that they have benefitted due to their race. In Studies 2a, 2b, and 2c, we examined whether people are more likely to recognize positive bias and take action to correct it in interpersonal interactions, even at a cost to themselves. In Study 3, we examined whether an indirect mention of a victim (through an emphasis on the zero-sum nature of a situation) has a similar effect as the more direct mention of a victim (through a victim anecdote) on recognizing positive bias and speaking up about it. Finally, in Studies 4a and 4b, we examined why salient victims lead to recognition of positive bias and corrective action. Specifically, and following TDM, we examined how a norm violation, negative affect, and obvious harm all play a role in raising awareness of positive bias and triggering action to correct it. Overall, understanding how to get people to recognize and act when they experience positive bias can help make societies fairer.

Transparency and Openness

We pre-registered the sample sizes, hypotheses, analyses, and exclusion processes for all studies reported in this paper, and we note a deviation to one pre-registration in a footnote (see SOM for additional pre-registered analyses). We aimed to have about 100 participants per cell,

following recent recommendations (Brysbaert, 2019), and we did not analyze our data until after data collection had been completed. We report all manipulations, measures, and exclusions for each study. All study materials and data are available at: https://osf.io/rmhw7/?view_only=3e3d1ba6b0924c30a5694400856fac53.

Study 1

In this study, we tested the effect of a salient victim on recognizing positive bias in one's societal treatment. Specifically, we were interested in whether White individuals would be more likely to report that they had benefitted from bias after reading about victims of potential racial profiling in traffic, retail, and travel scenarios.

Method

This study was pre-registered at: https://aspredicted.org/QK6_RPS. We recruited 875 White participants from the United States (439 men, 422 women, 14 other; M_{age} = 39.5, SD_{age} = 13.5) via Prolific. Despite the filter, 13 participants did not identify as White in our demographic survey, and they were excluded from the analyses. Because we had a secondary interest in examining moderation by political orientation, we chose this sample size to give us adequate power to detect an average-sized interaction in our field (see Aguinis et al. 2005).

All participants read short descriptions of the duties of traffic police, retail store employees, and TSA workers (e.g., "Traffic cops are tasked with stopping cars that are speeding or are otherwise breaking the law"). In the *control condition*, participants were not provided with any additional information. In the *salient victim condition*, participants were provided with an anecdote about a salient victim of what may have been racial profiling in each of the three situations. These anecdotes were pulled from online blurbs or from personal communications with victims. For example, for the traffic police anecdote, participants read the following:

A Black man named Robert was pulled over for speeding despite not driving over the speed limit. The state trooper who pulled him over then searched his car for drugs even though Robert was not doing anything suspicious. Robert had to stand out in the rain while his car was searched. The state trooper did not find anything illegal.

After each of the three blurbs, participants were asked to rate on a scale from 1 (*not at all*) to 7 (*very much*): "To what extent do you think you have personally benefited because of your race in terms of how you are treated when driving [shopping] [traveling]?" We took the mean of participants' three responses as our measure of bias recognition ($\alpha = .92$).

Finally, participants responded to several questions about themselves. Given that moral judgments can vary by one's political orientation (Graham et al., 2011; Van de Vyver et al., 2016), we also wanted to examine whether the salient victim effect is stronger among liberals than among conservatives in this context. Thus, we included a question where participants rated their political orientation on a scale from 1 (*very liberal*) to 11 (*very conservative*). Our sample leaned somewhat liberal ($M = 4.14$, $SD = 2.96$).

Participants also reported the frequency with which they drive, shop in retail stores, and travel in airports (*never* to *very often*). In line with our pre-registered plan, we excluded participants who said that they never drive, shop in retail stores, or fly from our analyses (as they would not have the opportunity to experience positive bias in these domains), leaving a final sample of 697 participants. A sensitivity analysis revealed that we had at least 80% power to detect a salient victim effect of $d = 0.21$ or larger.

Results and Discussion

As predicted, White participants in the salient victim condition were more likely to recognize the positive bias they had experienced in their own lives ($M = 4.91$, $SD = 1.80$),

compared to participants in the control condition ($M = 4.31$, $SD = 1.88$), $t(695) = 4.26$, $p < .001$, $d = 0.32$, 95% $CI_d = [0.17, 0.47]$. To determine whether the condition effect was moderated by political orientation, we ran a regression predicting bias recognition from condition (-1 = *control*, 1 = *salient victim*), political conservatism (centered), and their interaction (see Table 1). Political conservatism was negatively related to bias recognition in this context of race-based bias, but there was no evidence that political orientation moderated the salient victim effect. Thus, the size of the salient victim effect did not seem to change based on the political orientation of the participants. Because this study had a limited number of conservative participants ($N = 144$), we report a meta-analysis in SOM of conservatives across three separate studies ($N = 287$). The meta-analysis revealed that there is a significant salient victim effect for conservative participants, $Z = 3.17$, $p < .001$, $d = 0.38$, 95% $CI_d = [0.15, 0.62]$, supporting the idea that the salient victim effect occurs for people who hold different political views.

Table 1

Coefficients for a Linear Regression Model Predicting Bias Recognition in Study 1

	b [95% CI]	*t*	*p*	Partial *r*
Constant	4.61 [4.50, 4.72]	79.98	.000	
Salient victim v. control	0.32 [0.20, 0.43]	5.50	.000	.204
Political conservatism (centered)	-0.35 [-0.39, -0.31]	-17.96	.000	-.563
Victim*conservatism interaction	0.01 [-0.03, 0.04]	0.33	.743	.012

Notes. $N = 697$.

In sum, this study provides initial evidence for the salient victim effect. White participants recognized that they might have benefited from positive bias more when there were salient victims of racial profiling mentioned, and this effect seemed to occur for people of varying political views. Although these results may have been enhanced due to the potential

demand characteristics of our design, prior research has shown that it is not always easy to get White individuals to acknowledge their privilege (e.g., Phillips & Lowery, 2018); therefore, we are encouraged by these findings. While this study focused only on recognizing positive bias, the next studies will look at whether participants also take action to correct positive bias when presented with a salient victim.

Studies 2a, 2b, and 2c

In our next three studies, we tested whether the salient victim effect replicates when people experience positive bias in the moment. We further examined the important question of whether a salient victim leads people to act against such bias. It is often difficult for people to speak up or take other action against bias for various reasons, such as pressures to be polite or concerns about how the perpetrator will react (e.g., Czopp & Monteith, 2003; Rehg et al., 2008; Swim & Hyers, 1999). In the case of positive bias, it can be especially hard for the recipients to say or do anything about it, as doing so might cause them to lose out on their positive outcome or harm the person who tried to "help" them. We set up the following three studies to simulate this dilemma: Participants were given more than their fair share of something desirable due to their similarity with a decision maker, and we examined how they reacted. Past research has indicated that people receive advantages from those who have similar demographic characteristics and interests (Phillips et al., 2022). Thus, this seemed like an important type of positive bias to study.

Study 2a: Undeserved Lottery Tickets

Method

This study was pre-registered at: https://aspredicted.org/blind.php?x=ej4pc6. We recruited 250 participants from the United States (119 men, 125 women, 6 other; M_{age} = 31.3, SD_{age} = 11.6) via Prolific. As a cover story, we told participants we were interested in

"examining the teacher-student relationship and how it influences performance in a word game." Then, participants were ostensibly matched up virtually with two other Prolific users (who were not actually there). One supposed group member was assigned to the teacher role, and the other two people (including the real participant) were assigned to the student role. Participants were asked to answer some get-to-know-you and would-you-rather questions that would be shared with their teacher (e.g., "What is your birth month?" and questions from West et al., 2014, such as "Would you rather be extremely lucky or extremely smart?"). Afterwards, participants saw their teacher's responses, which were programmed to be very similar to their own.

Participants then completed a word game where they were told to find as many words as possible in a Boggle-style letter matrix in one minute. Participants learned that their teacher had been instructed to give out 20 lottery tickets across the two students based on how well the students performed on the word game. These lottery tickets would give participants a chance to earn a financial bonus.

After participants finished the word game, they waited about 20 seconds while their teacher ostensibly graded their work. They then received a message from their teacher stating that their initial score had been increased by five points, along with a justification indicating that the participant was the recipient of positive bias. In the *control condition*, the teacher told the participant (with "X" representing the actual number of words they listed): "You listed X words, but I'll bump your score up to a X+5. I don't think I'm supposed to do this, but you seem cool and we had a lot of similar answers on teh [*sic*] would you rather questions. Fun!" In the *salient victim condition*, the teacher's message started off the same, but then the teacher said: "The other student I'm grading seems lame. They said their favorite hobby is ant farms. So I'll just take some tickets from them, so you can have more." In both conditions, the teacher ended with "I

hope you win the lottery!" to remind participants of the connection between their score and the lottery tickets. We intentionally included informal writing and typos to enhance believability.

Participants were then given an opportunity to send a message back to their teacher in an open-ended text box (otherwise they could type N/A). After this, participants were asked to respond to 10 items about the word game and their teacher (1 = *strongly disagree*, 7 = *strongly agree*). Most were fillers, but three of these items made up our bias recognition scale (all reverse-scored; $\alpha = .84$): "The scoring process worked well," "The teacher's interaction with me was appropriate," and "The lottery tickets were allocated fairly." Then participants were asked if they would like to send any comments or concerns about the study, the word game, or their interaction with the teacher to the researcher through another text box. Finally, participants reported demographic information and were asked to guess the purpose of the study in a final text box.

To assess participants' willingness to speak up about the bias, we asked a research assistant blind to condition and hypothesis to code participants' first two open-ended responses (the message to the teacher and the message to the researcher) for whether the participants questioned or complained about the score increase or directly asked the teacher not to bump up their score. Thus, participants were given a score of a '1' on this variable if they spoke up to either the teacher and/or the researcher (e.g., "Hey Taylor, I appreciate the help, but I would prefer to be scored based on performance rather than similarity in personality," and "Thanks, but you don't have to. I don't want to take points from the other person."), and a '0' if they spoke up to neither party.

We also asked the research assistant to code participants' open-ended responses for suspicion (details in SOM). As pre-registered, we excluded data from participants who were

coded as suspicious, leaving a final sample size of 209. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $d = 0.39$ and $\varphi = .14$ or larger for each of our two dependent variables.

Results

As predicted, participants in the salient victim condition were more likely to recognize bias ($M = 4.33$, $SD = 1.59$), compared to participants in the control condition ($M = 3.43$, $SD = 1.46$), $t(207) = 4.26$, $p < .001$, $d = 0.59$, 95% $CI_d = [0.31, 0.87]$. Also as predicted, participants in the salient victim condition were more likely to speak up about the bias (25.3%), compared to participants in the control condition (11.8%), $\chi^2 = 6.32$, $p = .012$, $\varphi = .17$.

We then tested for mediation of the effect of the salient victim on speaking up through bias recognition (see Table 2 for the regression coefficients). According to the bootstrapping method for mediation with 10,000 iterations (Hayes, 2017), bias recognition was indeed a mediator (indirect effect = 0.62, 95% CI = [0.29, 1.14]). Thus, participants' higher likelihood of speaking up about the bias after being presented with a salient victim was explained by their greater likelihood of recognizing the bias.

Table 2

Coefficients for Logistic Regression Models Predicting Speaking Up in Study 2a

	b [95% CI]	*Z*	*p*	*b* [95% CI]	*Z*	*p*
Constant	-2.93 [-1.69, -4.18]	4.63	.000	-5.16 [-6.90, -3.41]	-5.78	.000
Salient victim v. control	0.92 [0.19, 1.66]	2.46	.014	0.40 [-0.41, 1.21]	0.97	.334
Bias recognition				0.69 [0.39, 0.99]	4.54	.000

Note. $N = 209$

Study 2b: Biased Offer in an Ultimatum Game

Method

This study was pre-registered at: https://aspredicted.org/1F1_JGN. We recruited 301 participants from the United States (151 men, 150 women, M_{age} = 41.1, SD_{age} = 13.1) via Connect, an online platform offered by CloudResearch. As a cover story, we told participants we were "interested in how and why people make decisions in an 'ultimatum game,' both in terms of their own proposals and whether they accept or reject other people's proposals." We then explained what an ultimatum game is. Like in Study 2a, we had participants answer questions about themselves (e.g., their demographic information, their hobbies, etc.) that would ostensibly be sent to an ultimatum game decision maker. Participants were then presented with an allocation decision that they had to accept or reject. They were told that the decision maker had to divide a 50-cent bonus between two other participants either randomly, evenly, or based on the participants' characteristics. This decision maker had chosen to allocate 40 out of the 50 cents to the participant based on their characteristics.

In the *control condition*, the decision maker sent a message that said, "I am [X age] and I want to help out someone who is of my same generation!" (We set it so that the decision maker was two years older than the participant to make them seem around the same age.) In the *salient victim condition*, the decision maker said the same thing but added, "I gave less to the person who was a different age." We also reminded participants in this condition in a couple other places that the other recipient would get a smaller amount (see more on our OSF page).

Participants were then asked if they would accept or reject the proposed bonus allocation (our measure of corrective action). If participants accepted the proposal, they would get 40 out of the 50 cents. If participants rejected the proposal, the bonus allocations would be made

randomly, and the decision maker would be penalized.[1] After making this decision, participants were asked to rate how much they agreed with various statements about why they decided to accept or reject the proposal (1 = *strongly disagree*, 7 = *strongly agree*). Three of these items made up our bias recognition scale (all reversed; $\alpha = .88$): "The decision maker made an appropriate decision given the available information," "The decision maker had good reasons for their allocation," and "The decision seemed fair overall."

Finally, participants were provided with two open-ended text boxes that asked if they had any comments about the study and if they had a guess about the purpose of the study. As pre-registered, we excluded data from participants who were coded as suspicious, leaving a final sample size of 297. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $d = 0.33$ and $\varphi = .11$ or larger for each of our two dependent variables.

Results

As predicted, participants in the salient victim condition were more likely to recognize bias ($M = 3.44$, $SD = 1.47$), compared to participants in the control condition ($M = 2.87$, $SD = 1.39$), $t(295) = 3.45$, $p < .001$, $d = 0.40$, 95% $CI_d = [0.17, 0.63]$. Also as predicted, participants in the salient victim condition were more likely to reject the proposal (13.4%), compared to participants in the control condition (3.4%), $\chi^2 = 9.72$, $p = .002$, $\varphi = .18$.

We then tested for mediation of the effect of the salient victim on rejecting the proposal through bias recognition (see Table 3 for the regression coefficients). According to the bootstrapping method for mediation with 10,000 iterations (Hayes, 2017), bias recognition was indeed a mediator (indirect effect = 0.57, 95% CI = [0.21, 1.10]). Thus, participants' higher

[1] We included this part to try to simulate the fact that taking action against positive bias can sometimes harm the person who displayed the positive bias. Some participants mentioned in their comments that they would have rejected the allocation if they were not going to harm the decision maker.

likelihood of rejecting the biased proposal after being presented with a salient victim was explained by their greater likelihood of recognizing the bias.

Table 3

Coefficients for Logistic Regression Models Predicting Rejecting the Proposal in Study 2b

	b [95% CI]	*Z*	*p*	*b* [95% CI]	*Z*	*p*
Constant	-3.35 [-4.24, -2.46]	7.37	.000	-7.17 [-9.10, -5.25]	-7.31	.000
Salient victim v. control	1.49 [0.48, 2.50]	2.89	.004	1.20 [0.10, 2.31]	2.14	.033
Bias recognition				0.99 [0.64, 1.35]	5.55	.000

Note. $N = 297$

Study 2c: Biased Task Allocation

Method

This study was pre-registered at: https://aspredicted.org/NSW_4F5. We recruited 194 undergraduates from a large university in the Northeastern United States (80 men, 112 women, 2 other; $M_{age} = 19.1$, $SD_{age} = 0.92$) who completed this study for extra credit in their class. As a cover story, we told participants that we were interested in "how people assign tasks to others, and then how people perform on those tasks depending on what tasks they were assigned." Participants started by rating how interesting various tasks sounded to them. Then they were told that they would be matched up with two other participants in the lab room, and of the three, one would be the decision maker and two would be recipients. In reality, all participants were assigned to the recipient role, and as in our previous two studies, all interactions across "participants" were pre-planned. To make the groups seem realistic, we had participants take the study at the same time as other people. There were a few time slots where a single participant took the study, in which case the experimenters were instructed to make it seem like there were other people in the lab room (as participants could not see everyone else).

Participants then answered questions about themselves, like those in Study 2a, that would ostensibly be sent to their decision maker, who then had to allocate experimental tasks across the two recipients. Participants learned that the decision maker had to choose how to allocate a humor perceptions task (4.7/5 stars), a word game task (4.8/5 stars), a letter counting task (2.1/5 stars), and a safety video task (1.8/5 stars), and that their decision maker gave them the two higher rated tasks. They then received a message from the decision maker. In the *control condition*, the decision maker said, "I chose to give you the two fun tasks becuase [*sic*] we had a lot of similar answers on the get to know you quesions [*sic*]!" In the *salient victim condition*, the decision maker said the same thing except also added, "The other person will get the two boring tasks." There were also additional mentions of the victim in this condition.

For our measure of corrective action, participants were then asked if they would like to accept the allocation and do both higher rated tasks, or if they would like to reject the allocation and do one of the higher rated tasks and one of the lower rated tasks. If they rejected the offer, their decision maker would have to do both lower rated tasks. To avoid demand effects that were possibly present in our previous studies, we did not include a bias recognition measure in this study.

Participants then went on to complete either both fun tasks or one fun task and one boring task (depending on whether they accepted or rejected the proposal) and some demographic items. Finally, in an open-ended text box, participants were asked to report whether they had a guess about the purpose of the study and if anything seemed strange to them. As pre-registered, we excluded data from participants who were coded as suspicious based on this text box, leaving a final sample size of 178. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $\varphi = .15$ or larger for the effect of condition on corrective action.

Results

As predicted, participants in the salient victim condition were more likely to reject the biased allocation (18.2%), compared to participants in the control condition (6.7%), $\chi^2 = 5.45$, $p = .020$, $\varphi = .17$.

Discussion

In these studies, a salient victim led participants to be more likely to recognize that they benefited from bias (Studies 2a–2b) and take action to correct the bias (Studies 2a–2c). In the control condition, even though the decision maker indicated that participants were getting a good outcome that was more than their fair share (more lottery tickets than earned, more than 50% of a shared bonus, or more than 50% of the fun tasks), participants usually did not protest. However, when the victim who would suffer from the bias was highlighted, participants were more likely to act against the positive bias, through speaking up or rejecting the proposal. Participants may have felt more compelled to act once they were aware of the unfair positive bias (Jones, 1991), as not acting might have made them feel like a moral hypocrite (Batson & Thompson, 2001). Participants were more likely to correct the positive bias in the salient victim condition even though doing so would be at a cost to themselves (i.e., fewer lottery tickets, a potentially worse bonus allocation, and fewer fun tasks) and possibly also their decision maker.

Although we found promising results in these behavioral studies, there were also limitations. While we assume that people need to recognize positive bias before they will correct it, and our mediation results were consistent with this, it is important to note that mediation analyses are limited (Bullock et al., 2010), and we cannot be sure of the causal order between bias recognition and corrective action. In Studies 2a and 2b, we asked the bias recognition questions (mediator) *after* participants had the opportunity to take some corrective action (DV)

because we did not want to influence their behavior by hinting at the unfairness of the bias through our questions. Also, while we tried to make the interactions seem as realistic as possible, some participants were still suspicious that they were fabricated and were therefore excluded from the analyses presented above. As pre-registered, we also analyzed the data without any exclusions for suspicion, and we obtained the same patterns of results (see SOM).

Study 3

In Studies 1–2c, we manipulated the victim in a direct way. In Study 3, we examined whether an indirect manipulation of a victim—where participants are reminded of the limited number of people who can benefit (i.e., an emphasis on the zero-sum nature of the situation)—has a similar effect on participants' likelihood of recognizing bias and speaking up about it.

Method

This study was pre-registered at: https://aspredicted.org/blind.php?x=848iw6. We recruited 405 participants from the United States (214 men, 191 women; M_{age} = 39.8, SD_{age} = 12.68) from CloudResearch's approved mTurk participant list. Participants were first asked to imagine that they had worked at a company for six months and were the second-best salesperson. This company gives out bonuses for outstanding performance, but "only employees who have worked at the company for at least a year are eligible for the bonus." Participants also found out that their cousin oversees the committee that makes the decisions about the bonuses. Despite their ineligibility due to the tenure rule for bonuses, participants learned they were selected for a bonus anyway. Thus, the scenario implied that the cousin may have violated the tenure rule on behalf of the participant because they were family (i.e., positive bias via nepotism). The additional information participants received varied by condition according to a 2 (direct mention: salient victim v. not) x 2 (indirect mention: zero-sum v. not) between-subjects design.

For the direct mention of the victim, we manipulated whether a specific victim was emphasized or not, like in previous studies. Half of the participants received no information about the other people who did or did not receive a bonus (the *control condition*). The other half (the *salient victim condition*) read:

> Your cousin also tells you who the other winners are. You realize that your senior colleague who has been working at the company for eight years and who is the third best salesperson in the company will not be one of the five people getting the bonus this year**.** This senior colleague did not receive the bonus last year either despite excellent performance then.

For the indirect mention of the victim, we manipulated whether there were a limited number of bonuses or not (i.e., whether it was a zero-sum situation). In the *non-zero-sum condition*, participants read: "The company will give out bonuses to however many employees meet the criteria." In the *zero-sum condition*, participants read: "The company will give out five total bonuses to the top five employees who meet the criteria."

Participants then rated the extent to which they agreed with the following statements (1 = *strongly disagree*, 7 = *strongly agree*): "The decision feels unfair," "The decision to give me the bonus has negative consequences," "I deserved the bonus (reversed)," "The committee was justified in giving me the bonus (reversed)," "It feels wrong to accept the bonus," "The committee made a biased choice," and "I received the bonus because of favoritism." We took the mean of these seven items as a measure of bias recognition ($\alpha = .90$). Participants then responded to a 5-item scale of speaking up about the bias ($\alpha = .89$; e.g., "I would tell the committee that they did not have to break the rule for me;" see all items on our OSF page).

At the end of the study, participants answered demographic items and a memory attention check asking who assigned the bonuses (i.e., the cousin). As pre-registered, participants who incorrectly answered the memory question were excluded, leaving a final sample of 384 participants. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $\eta^2 = .018$ or larger for each of our two dependent variables.

Results and Discussion

Results of a 2 x 2 factorial ANOVA on bias recognition revealed a significant main effect of the direct mention of the victim, $F(1, 380) = 44.19$, $p < .001$, $\eta^2 = .10$: Participants were more likely to recognize bias when they read about the person who did not receive the bonus (the salient victim) than when they did not read about this person. There was also a significant main effect of the indirect mention of the victim, $F(1, 380) = 12.86$, $p < .001$, $\eta^2 = .03$: The bonus decision was rated as more biased when there was a zero-sum situation as opposed to when there was not. There was also a significant interaction, $F(1, 380) = 5.09$, $p = .025$, $\eta^2 = .01$ (see Figure 1). The direct mention of the victim led to a greater likelihood of recognizing bias, both in the zero-sum situation, $t(380) = 3.15$, $p = .002$, $d = 0.45$, 95% $CI_d = [0.16, 0.73]$, and in the non-zero-sum situation, $t(380) = 6.20$, $p < .001$, $d = 0.91$, 95% $CI_d = [0.61, 1.21]$; however, the salient victim effect was larger in the non-zero-sum situation.

Results of a 2 x 2 factorial ANOVA on speaking up revealed a significant main effect of the direct mention of the victim, $F(1, 380) = 6.38$, $p = .012$, $\eta^2 = .02$: Participants said they were more likely to speak up when they learned of the salient victim. There was also a significant main effect of the indirect mention of the victim, $F(1, 380) = 7.31$, $p = .007$, $\eta^2 = .02$: Participants said they would speak up more in the zero-sum situation. There was also a marginally significant interaction, $F(1, 380) = 3.67$, $p = .056$, $\eta^2 = .01$ (see Figure 2). The direct

mention of the victim did not lead to a greater likelihood of speaking up in the zero-sum situation, $t(380) = 0.44$, $p = .661$, $d = 0.06$, 95% $CI_d = [-0.22, 0.34]$, but it did in the non-zero-sum situation, $t(380) = 3.09$, $p = .002$, $d = 0.45$, 95% $CI_d = [0.16, 0.74]$.

Figure 1

Means and 95% CIs for Bias Recognition per Condition in Study 3

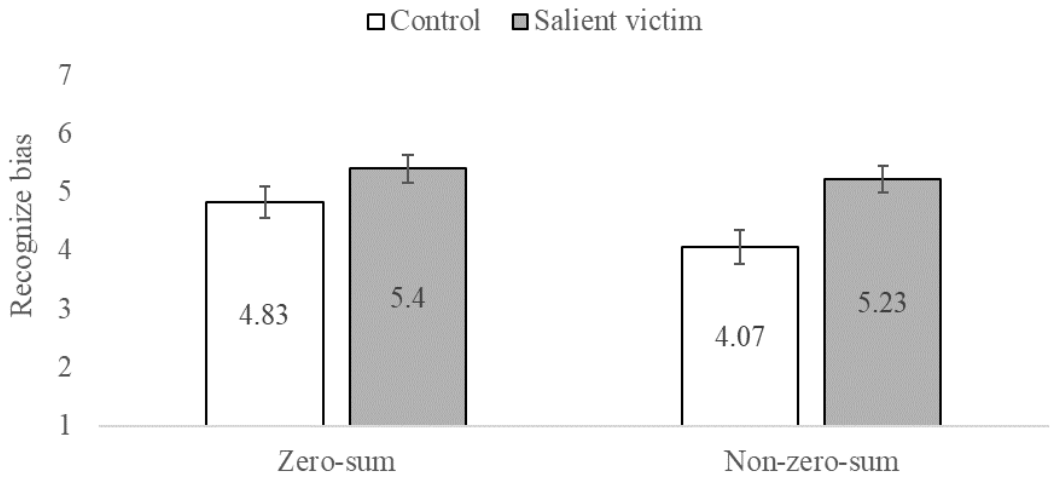
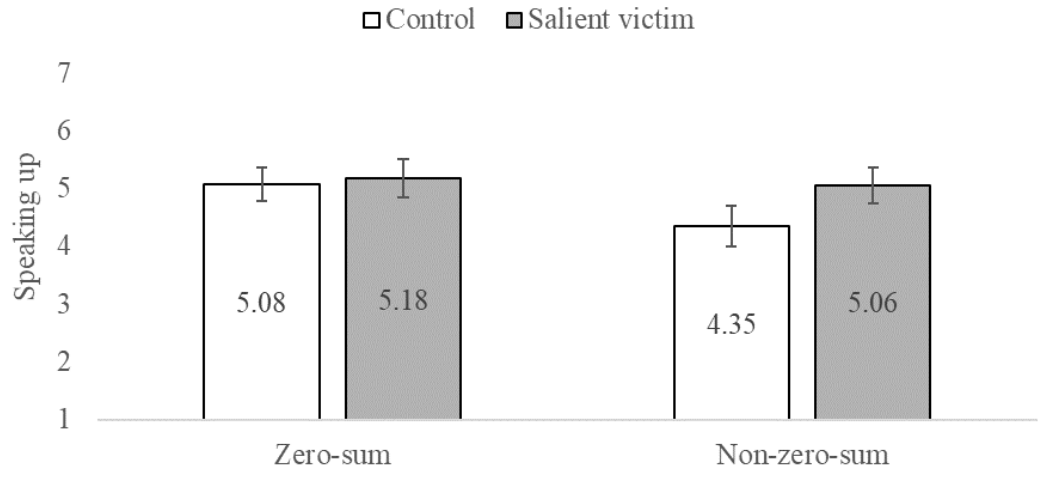


Figure 2

Means and 95% CIs for Speaking Up per Condition in Study 3



We then conducted a mediation analysis to examine whether the two main effects found for speaking up (i.e., the direct mention of the victim and the indirect mention of the victim)

were explained by greater bias recognition (see Table 4 for the regression coefficients). According to the bootstrapping method with 10,000 iterations, bias recognition mediated the relationship between the direct mention of the victim and speaking up (indirect effect = 0.29, 95% CI = [0.20, 0.40]). Bias recognition also mediated the relationship between the indirect mention of the victim and speaking up (indirect effect = 0.16, 95% CI = [0.07, 0.25]).[2]

Table 4

Coefficients for Linear Regression Models Predicting Speaking Up in Study 3

	b [95% CI]	*t*	*p*	*b* [95% CI]	*t*	*p*
Constant	4.92 [4.76, 5.07]	61.91	.000	1.59 [1.08, 2.11]	6.07	.000
Salient victim v. control	0.20 [0.04, 0.36]	2.53	.012	-0.09 [-0.23, 0.04]	-1.35	.179
Zero-sum v. not	0.21 [0.06, 0.37]	2.70	.007	0.06 [-0.08, 0.19]	0.83	.405
Salient victim X zero-sum	-0.15 [-0.31, 0.00]	-1.91	.056	-0.05 [-0.18, 0.08]	-0.78	.433
Bias recognition				0.68 [0.58, 0.78]	13.08	.000

Notes. *N* = 384. Coding: salient victim =1, control = -1; zero-sum = 1, non-zero-sum = -1

Overall, Study 3 showed that people were more likely to recognize positive bias if there was either a specific victim mentioned (the direct victim manipulation) or if the zero-sum nature of the situation was emphasized (the indirect victim manipulation), and this heightened bias recognition seemed to account for participants' greater likelihood of speaking up after both types of victim indications.[3] Thus, there are multiple ways of producing the salient victim effect. While the direct victim manipulation relied on giving participants detailed information about the victim (as opposed to no information), the results from the indirect manipulation were similar when we simply told participants that there were limited bonuses (as opposed to no limit). This suggests

[2] There were some differences between the pre-registration and what we ended up focusing on in our write-up. Please see SOM for the other pre-registered analyses.

[3] Although we believe that people need to recognize positive bias before they can act against it, we again note that mediation analyses are limited (Bullock et al., 2010), and we cannot be sure about the causal order of our variables.

that the specific information given about the victim likely does not drive our effects. In short, drawing attention to the harm in either way (a victim anecdote or a zero-sum emphasis) seemed to remove the blinders and compel people to act against positive bias, whereas having neither led to the least bias recognition and to the least speaking up about the bias.

Studies 4a and 4b

In our final studies, we wanted to examine why the salient victim effect occurs. We have argued that a salient victim increases awareness of harm, demonstrates that a norm violation took place, and produces negative affect, which, following TDM, would lead people to recognize and condemn the positive bias (and then subsequently take corrective action). To test if this is indeed the case, we examined bias recognition and corrective action after exposure to a salient victim in special circumstances when these factors were reduced. Specifically, we asked participants to imagine that they had been given a promotion they were not ready for by their father (positive bias via nepotism), and we examined whether people were as likely to recognize the bias and reject the undeserved promotion when there was a victim who experienced less harm (Study 4a), when there was less evidence of a norm violation (Study 4b), or when negative affect was reduced (Study 4b).

Study 4a

Method

This study was pre-registered at: https://aspredicted.org/9TQ_HXF. We recruited 302 participants from the United States (152 men, 150 women; M_{age} = 37.88, SD_{age} = 12.14) from Connect. We asked participants to imagine that they work at a company where their father is the boss. Their father has to retire earlier than expected and wants to promote them to fill his position. To demonstrate further that there was bias involved, we told participants, “You were

not expecting to move up this quickly, and you aren't sure if you are ready. However, your dad tells you that he is confident that you will do a good job." What participants read next varied by condition. In the *control condition*, participants were not given any other information. In the *salient victim condition*, participants read, "You realize that your colleague Taylor, who has been working at the company longer than you and who is a very good employee, won't be getting the promotion." In the *salient victim with reduced harm condition*, they read the same thing plus, "However, you also know that Taylor was recently offered an appealing job at a different company and therefore has another option."

Participants then responded (1 = *strongly disagree*, 7 = *strongly agree*) to a question asking, "How likely are you to accept the promotion?" and the 7-item bias recognition scale used in Study 3 but with minor edits to fit the context of this study ($\alpha = .89$). At the end of the study, participants answered demographic items and a memory attention check asking who offered them the promotion. As pre-registered, participants who incorrectly answered the memory question were excluded, leaving a final sample of 293 participants. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $d = .33$ or larger for each of our two dependent variables.

Results

The results of a one-way ANOVA showed differences across conditions in bias recognition, $F(2, 290) = 6.78$, $p = .001$, $\eta^2 = .05$, and accepting the promotion, $F(2, 290) = 11.47$, $p < .001$, $\eta^2 = .07$. As predicted, participants in the salient victim condition were more likely to recognize bias ($M = 4.48$, $SD = 1.10$), compared to participants in the control condition ($M = 3.90$, $SD = 1.22$), $t(290) = 3.44$, $p < .001$, $d = .49$, 95% $CI_d = [0.21, 0.77]$, and to participants in the reduced harm condition ($M = 3.99$, $SD = 1.22$), $t(290) = 2.85$, $p = .005$, $d = .41$, 95% $CI_d =$

[0.13, 0.69]. Also as predicted, participants in the salient victim condition were less likely to accept the promotion (M = 5.10, SD = 1.68), compared to participants in the control condition (M = 5.90, SD = 1.18), $t(290) = 4.09$, $p < .001$, $d = 0.58$, 95% CI_d = [0.30, 0.87], and to participants in the reduced harm condition (M = 5.93, SD = 1.18), $t(290) = 4.20$, $p < .001$, $d = 0.60$, 95% CI_d = [0.32, 0.89].

We then examined whether the lower likelihood of accepting the promotion in the salient victim condition (v. the other two conditions) was explained by greater bias recognition using the bootstrapping method for mediation with 10,000 iterations (see Table 5 for the regression coefficients). Indeed, bias recognition mediated the effect when comparing the salient victim condition to the control condition (indirect effect = 0.29, 95% CI = [0.13, 0.46]), and to the reduced harm condition (indirect effect = 0.24, 95% CI = [0.08, 0.42]).

Table 5

Coefficients for Linear Regression Models Predicting Accepting the Promotion in Study 4a

	b [95% CI]	*t*	*p*	*b* [95% CI]	*t*	*p*
Constant	5.10 [4.83, 5.37]	36.93	.000	7.33 [6.74, 7.93]	24.25	.000
Control (1) v. salient victim (0)	0.80 [0.41, 1.18]	4.09	.000	0.51 [0.15, 0.86]	2.83	.005
Reduced harm (1) v. salient victim (0)	0.83 [0.44, 1.21]	4.20	.000	0.58 [0.23, 0.94]	3.24	.001
Bias recognition				-0.50 [-0.62, -0.38]	-8.10	.000

Notes. N = 293

Study 4b

Method

This study was pre-registered at: https://aspredicted.org/LFY_S37. We recruited 400 participants from the United States (204 men, 195 women, 1 other; M_{age} = 39.7, SD_{age} = 11.89) from Connect. Participants were asked to read and imagine that they were in the same scenario

from Study 4a (i.e., participants read that they were offered a promotion by their father). We then randomly assigned participants to one of four conditions.[4]

The *control condition* and *salient victim condition* were the same as in Study 4a. In the *salient victim with reduced norm violation condition*, participants read everything from the salient victim condition plus a final comment that said, "However, this is a family firm, and it is common in this company and similar others to prioritize family members." In the *salient victim with reduced negative affect condition*, as an add-on to the salient victim condition, participants instead read, "However, you're not even sure if Taylor would like doing this new job anyway, so you don't feel bad."

Participants then answered the exact same DVs and other questions as in Study 4a. As pre-registered, participants who incorrectly answered the memory question were excluded, leaving a final sample of 388 participants. A sensitivity analysis revealed that we had at least 80% power to detect effect sizes of $d = 0.29$ or larger for each of our two dependent variables.

Results

The results of a one-way ANOVA showed differences across conditions in bias recognition, $F(3, 384) = 2.94$, $p = .033$, $\eta^2 = .02$, and accepting the promotion, $F(3, 384) = 14.94$, $p < .001$, $\eta^2 = .11$. As predicted, participants in the salient victim condition were more likely to recognize bias ($M = 4.30$, $SD = 1.14$), compared to participants in the control condition ($M = 3.89$, $SD = 1.06$), $t(384) = 2.63$, $p = .009$, $d = .38$, 95% $CI_d = [0.09, 0.66]$, to participants in the reduced negative affect condition ($M = 3.94$, $SD = 1.03$), $t(384) = 2.26$, $p = .025$, $d = .32$, 95% $CI_d = [0.04, 0.61]$, and to participants in the reduced norm violation condition ($M = 3.93$, $SD = 1.15$), $t(384) = 2.34$, $p = .020$, $d = .34$, 95% $CI_d = [0.05, 0.62]$. Also as predicted, participants in

[4] See SOM for the results of our manipulation checks.

the salient victim condition were less likely to accept the promotion (M = 5.19, SD = 1.47), compared to participants in the control condition (M = 6.13, SD = 1.02), $t(384) = 5.80$, $p < .001$, $d = 0.83$, 95% CI_d = [0.54, 1.12], to participants in the reduced negative affect condition (M = 5.96, SD = 0.98), $t(384) = 4.72$, $p < .001$, $d = 0.68$, 95% CI_d = [0.39, 0.96], and to participants in the reduced norm violation condition (M = 6.11, SD = 1.00), $t(384) = 5.65$, $p < .001$, $d = 0.82$, 95% CI_d = [0.53, 1.10]. See Figure 3 for a visual display of these results.

Figure 3

Means and 95% CIs for Both DVs per Condition in Study 4

We then examined whether the lower likelihood of accepting the promotion in the salient victim condition (v. the other three conditions) was explained by greater bias recognition using the bootstrapping method for mediation with 10,000 iterations (see Table 6 for the regression coefficients). Indeed, bias recognition mediated the effect when comparing the salient victim condition to the control condition (indirect effect = 0.18, 95% CI = [0.04, 0.34]), to the reduced

negative affect condition (indirect effect = 0.16, 95% CI = [0.02, 0.31]), and to the reduced norm violation condition (indirect effect = 0.16, 95% CI = [0.02, 0.32]).[5]

Table 6

Coefficients for Linear Regression Models Predicting Accepting the Promotion in Study 4b

	b [95% CI]	*t*	*p*	*b* [95% CI]	*t*	*p*
Constant	5.19 [4.96, 5.42]	44.74	.000	7.10 [6.64, 7.55]	30.67	.000
Control (1) v. salient victim (0)	0.94 [0.62, 1.26]	5.80	.000	0.76 [0.47, 1.05]	5.12	.000
Reduced norm violation (1) v. salient victim (0)	0.93 [0.60, 1.25]	5.65	.000	0.76 [0.47, 1.06]	5.11	.000
Reduced negative affect (1) v. salient victim (0)	0.77 [0.45, 1.09]	4.72	.000	0.61 [0.32, 0.91]	4.12	.000
Bias recognition				-0.44 [-0.54, -0.35]	-9.26	.000

Notes. $N = 388$.

Discussion

Studies 4a and 4b provide information about why a salient victim produces more bias recognition and corrective action. Consistent with TDM, it seems that a salient victim highlights the harm, produces negative affect in the beneficiaries, and shows that there was a norm violation. When the salient victim does not cause these reactions, people do not recognize positive bias or act against it to the same degree.

General Discussion

In a series of seven pre-registered studies, we found that people recognize and act against positive bias (i.e., receiving an unmerited favorable outcome due to preferential treatment) more when there is a salient victim. People who experience positive bias may not even realize that anything unfair has happened or that others may have been harmed by the treatment they benefited from (Pilot Study 1). However, a salient victim can break this tendency and lead

[5] See SOM for an additional study aiming to examine the role of negative affect and norms.

people to recognize when they are beneficiaries of bias (Studies 1–4b), and even encourage them speak up or take other action to correct the bias (Studies 2–4b). This suggests that when people accept positively biased treatment without complaint, they are not merely indifferent to the negative consequences; they might just not be sufficiently aware that they received biased treatment that harmed others.

The salient victim effect occurred when participants considered their own real-life positive outcomes (i.e., when they were potential beneficiaries of racial bias), in ostensible interpersonal interactions, and in hypothetical scenarios. The salient victim effect emerged for both direct and indirect mentions of the victim, and it was robust to different reasons for the positive bias (i.e., participants were White, they were similar to the decision maker, or they were related to the decision maker). Consistent with TDM, a salient victim seems to have its effects because it gets people to identify a norm violation, feel negative emotions, and notice harm. In sum, the salient victim has powerful effects across situations, even leading people to be more likely to give up positive outcomes they received due to positive bias. Although there is research showing that the way inequity is framed can affect how people react to it (e.g., Dietze & Craig, 2021; Jun et al., 2022; Lowery et al., 2009; Lowery et al., 2012; Rosette & Koval, 2018), we take a step back and examine how to get people to recognize that there was any inequity in the first place. This research adds to the growing bodies of literature on how victims can affect moral judgments and behaviors across a variety of contexts (e.g., Jenni, & Loewenstein, 1997; Jordan & Kouchaki, 2021; Small & Loewenstein, 2003).

Theoretical and Practical Contributions

Overall, our paper and our introduction of the salient victim effect make three main theoretical contributions. First, most research on justice and morality (e.g., Barclay et al., 2005;

Mikula et al., 1998; Zitek et al., 2010), bias awareness (e.g., Daumeyer et al., 2019; Perry et al., 2015), confronting bias (e.g., Burns & Monteith, 2019; Rattan & Dweck, 2010; Rattan & Dweck, 2018), and inequitable outcomes (e.g., Hebl et al., 2002; Milkman et al., 2015) has focused on the negative or disadvantaged side of inequity (e.g., recipients of undeserved negative events or perpetrators/victims of negative bias). Much less research has focused on the positive or advantaged side, even though both disadvantaging and advantaging mechanisms produce inequity (Phillips et al., 2022). We expand past research by examining one advantaging mechanism—positive bias—which is an understudied yet common form of bias in society.

Second, our research presents a way to help people recognize when a positive outcome is biased and unfair, which has sometimes proven difficult in past research due to people's tendency to justify their positive outcomes (e.g., Ellard & Bates, 1990). Although some researchers argue that privilege may be invisible to those who benefit from it (e.g., White individuals; Phillips & Lowery, 2018), we demonstrate that a salient victim can help make privilege more visible (see Study 1). Moreover, consistent with the idea that moral intensity and increased moral awareness are important for moral behavior (Jones, 1991), we also found that bias recognition predicts a greater likelihood of taking corrective action, even when the individual might suffer as a result. Our research thus reveals an important cue—a salient victim—that can help people recognize and act against the biases they benefit from, and potentially enhance social justice in the process.

Third, contrary to prior research on the pitfalls of zero-sum thinking (e.g., Davidai & Ongis, 2019; Norton & Sommers, 2011; Sirola & Pitesa, 2017, Wilkins et al., 2015), we demonstrate that there is a circumstance where zero-sum thinking can be beneficial. Specifically, in Study 3, participants were more likely to recognize positive bias and speak up about it when

we told them that bonuses were limited (i.e., when we emphasized that it was a zero-sum situation). Thus, positive bias is a case where it helps for people to realize that when one person wins, someone else loses.

From a practical perspective, small acts of positive bias are easy to ignore, but they can build up over time, producing toxic cultures (Sull et al., 2022). As demonstrated in our research, salient victims help beneficiaries recognize and act against positive bias, which can reduce harm and increase fairness in society. For example, at the height of the #BlackLivesMatter movement, after hearing victim anecdotes, many White celebrities and corporate leaders concluded that positive bias helped them get their jobs, and some even stepped down from their positions so that Black individuals could fill them (e.g., Hatmaker & Lunden, 2020; Zlotnick, 2020).

Beyond examples of rectifying positive bias by transferring specific opportunities from one person to another, there are many other possible advantages of having people recognize and act against positive bias. For example, an organization that minimizes the effects of positive bias might be seen as higher in organizational justice, which is associated with various positive consequences, including increased job performance and commitment among employees (e.g., Cropanzano et al., 2007). Moreover, individuals who speak up about the positive bias they have benefitted from may be trusted more, as they demonstrate their commitment to fairness by calling out potentially unethical behavior within their organizations (Kennedy & Schweitzer, 2018). Individuals who try to correct positive bias may also be perceived as allies to disadvantaged individuals, and as a result they may help organizations become more inclusive (Melaku et al., 2020).

Limitations and Future Directions

Notwithstanding the contributions of this research, our studies are not without limitations that could inspire future directions. Although we have focused on people's own outcomes, a failure to recognize positive bias may also emerge when people are judging the outcomes of others, or when they are making decisions about others. Indeed, Phillips and Jun (2022) showed that observers judge a biased action as less problematic when it is framed positively (focusing on the people who benefit rather than on those who lose); for example, people are less likely to perceive "favoring men" (v. "disfavoring women") as discrimination. Other scholars have discussed the bounded ethicality of decision makers like employers who engage in favoritism without realizing that they are being unethical (Banaji et al., 2003; Gino et al., 2010). Making a victim salient could induce unaffected third parties to act against positive bias or help decision makers avoid engaging in positive bias in the first place. Future research could examine how the salient victim effect applies in these other contexts.

An additional limitation is that we do not conclusively know whether people are more likely to act against positive bias after being exposed to a salient victim because they personally want to be moral (Mazar et al., 2008) or because they do not want others to judge them as immoral (Batson & Thompson, 2001; Lönnqvist et al., 2014). That is, does recognizing positive bias (due to salient victims) help people overcome moral hypocrisy because it increases the personal costs (i.e., moral self-view) or the public costs (i.e., public moral image) of hypocritical behavior? It is likely a combination of both, and future research should explore these paths further.

Moreover, we currently have a limited understanding of factors that undermine or enhance the salient victim effect. We found some initial evidence that the salient victim effect may exist across political orientations in the surprising case of positive bias due to race (see

Study 1 and the two replication studies in SOM). In an additional study (Supplemental Study 2 in SOM), we examined whether narcissism or moral disengagement moderated the salient victim effect, but neither did. Despite the lack of evidence for the moderators we have explored, we believe many are likely. For example, it is possible that participants from countries other than the United States would react differently to positive bias, as fairness perceptions tend to vary across national cultures (e.g., Schäfer et al., 2015). It is also possible that already disadvantaged groups, who may see themselves as victims, might be less likely to show the salient victim effect, as they might perceive that the benefits they received through positive bias are a fair way to make up for their past victimization (see also Zitek et al., 2010).

Relatedly, we would like to acknowledge that the rate of acting against positive bias in our studies was not very high. Although the salient victim caused people to be more likely to give up their personal benefits, many people still wanted to keep their lottery tickets, bonus, good tasks, and even their hypothetical job offer. Because self-interest can overpower moral integrity (Batson & Thompson, 2001), we imagine that it would be even harder to get people to give up a larger reward (e.g., a large bonus or a real job offer). Indeed, other research confirms that people might not be as likely to correct an inequity if they have more to lose from doing so: For example, NBA players will miss a free throw after an incorrect, beneficial foul call, but only if their team is already winning (Haynes & Gilovich, 2010). Future research should examine how to use the salient victim effect as a real-world intervention to get people to recognize and correct positive bias, and whether the effect can be strengthened so that people act against positive bias even when they have a lot more to lose.

Finally, information about victims might sometimes lead to a threat response, which may explain why social movements (e.g., #MeToo, #BlackLivesMatter) elicit support from some

individuals but resistance from others. For example, past research has shown that when White Americans read about the advantages they experience in society over Black Americans, they are *not* more likely to report that they believe in White privilege (Phillips & Lowery, 2015). Maybe a key element in our studies is that participants did not feel blamed for their advantages (see Fath et al., 2022, for more on threat and White privilege). More research is needed to address these and similar questions.

Conclusion

As a first step toward ensuring that our world is fair, it is imperative that people recognize and react when they are victims of bias *and* when they are beneficiaries of bias. Our research provides initial evidence that we are more likely to recognize when we benefit from positive bias, and to act against such bias, when a victim of that bias is made salient—i.e., the salient victim effect.

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 2*, (pp. 267−299). New York: Academic Press.

Aguinis, H., Beaty, J. C., Boik, R. J., Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*, 94–107. https://doi.org/10.1037/0021-9010.90.1.94

Austin, W., & Walster, E. (1975). Equity with the world: The trans-relational effects of equity and inequity. *Sociometry, 38*(4), 474–496. https://doi.org/10.2307/2786362

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *WIREs Cognitive Science, 4*, 169−178. https://doi.org/10.1002/wcs.1216

Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003). How (un)ethical are you? *Harvard Business Review, December 81*(12), 56–64.

Barclay, L. J., Skarlicki, D. P., & Pugh, S. D. (2005). Exploring the role of emotions in injustice perceptions and retaliation. *Journal of Applied Psychology, 90*(4), 629–643. https://doi.org/10.1037/0021-9010.90.4.629

Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General, 141*(4), 757–773. https://doi.org/10.1037/a0027588

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569–579. https://doi.org/10.1037/0022-3514.54.4.569

Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current Directions in Psychological Science, 10*(2), 54–57. https://doi.org/10.1111/1467-8721.00114

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77*(3), 525–537. https://doi.org/10.1037/0022-3514.77.3.525

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bohns, V. K. (2016). (Mis)Understanding our influence over others: A review of the underestimation-of-compliance effect. *Current Directions in Psychological Science, 25*(2), 199−123. https://doi.org/10.1177/0963721415628011

Brockner, J., Greenberg, J., Brockner, A., Bortz, J., Davy, J., & Carter, C. (1986). Layoffs, equity theory, and work performance: Further evidence of the impact of survivor guilt. *Academy of Management Journal, 29*(2), 373–384. https://doi.org/10.2307/256193

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*(16), 1–38. http://doi.org/10.5334/joc.72

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550–558. https://doi.org/10.1037/a0018933

Burns, M. D., & Monteith, M. J. (2019). Confronting stereotypic biases: Does internal versus external motivational framing matter? *Group Processes and Intergroup Relations, 22*(7), 930–946. https://doi.org/10.1177/1368430218798041

Cao, J., Kleiman-Weiner, M., & Banaji, M. R. (2019). People make the same Bayesian judgment they criticize in others. *Psychological Science, 30*(1), 20–31. https://doi.org/10.1177/095679761880575

Carlson, R. W., Adkins, C., Crockett, M. J., & Clark, M. S. (2022). Psychological selfishness. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916211045692

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In *Handbook of social cognition: Basic processes; Applications, Vols. 1-2, 2nd ed* (pp. 323–417). Lawrence Erlbaum Associates, Inc.

Cropanzano, R., Bowen, D. E., & Gilliland, S. W. (2007). The management of organizational justice. *Academy of Management Perspectives, 21*(4), 24–48. https://doi.org/10.5465/amp.2007.27895338

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin, 29*(4), 532–544. https://doi.org/10.1177/0146167202250923

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology, 84*. https://doi.org/10.1016/j.jesp.2019.04.010

Davidai, S., & Gilovich, T. (2016). The headwinds/tailwinds asymmetry: An availability bias in assessments of barriers and blessings. *Journal of Personality and Social Psychology, 111*(6), 835–851. https://doi.org/10.1037/pspa0000066

Davidai, S., & Ongis, M. (2019). The politics of zero-sum thinking: The relationship between political ideology and the belief that life is a zero-sum game. *Science Advances, 5*(12), aay3761. https://doi.org/10.1126/sciadv.aay3761

Dietze, P., & Craig, M. A. (2021). Framing economic inequality and policy as group disadvantages (versus group advantages) spurs support for action. *Nature Human Behaviour, 5*, 349–360. https://doi.org/10.1038/s41562-020-00988-4

Dion, K. L., & Earn, B. M. (1975). The phenomenology of being a target of prejudice. *Journal of Personality and Social Psychology, 32*(5), 944–950. https://doi.org/10.1037//0022-3514.32.5.944

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568–584. https://doi.org/10.1037/0022-3514.63.4.568

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *The psychology of learning and motivation* (pp. 307–338). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(08)00410-6

Dunning, D. (2007). Self-image motives and consumer behavior: How sacrosanct self-beliefs sway preferences in the marketplace. *Journal of Consumer Psychology*, *17*(4), 237–249. https://doi.org/10.1016/S1057-7408(07)70033-5

Ellard, J. H., & Bates, D. D. (1990). Evidence for the role of the justice motive in status generalization processes. *Social Justice Research, 4*(2), 115–134. http://dx.doi.org/*10.1007*/*BF01053242*

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

Fath, S., Ma, A., & Rosette, A. S. (2022). Self-views of disadvantage and success impact perceptions of privilege among White men. *Organizational Behavior and Human Decision Processes, 169.* https://doi.org/10.1016/j.obhdp.2021.104114

Feather, N. T., & McKee, I. R. (2009). Differentiating emotions in relation to deserved or undeserved outcomes: A retrospective study of real-life events. *Cognition and Emotion, 35*(5), 955–977. https://doi.org/10.1080/02699930802243378

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140. https://doi.org/10.1038/415137a

Gaucher, D., Hafer, C., Kay, A. C., & Davidenko, N. (2010). Compensatory rationalizations and the resolution of everyday undeserved outcomes. *Personality and Social Psychology Bulletin, 36*(1), 109–118. https://doi.org/10.1177/0146167209351701

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*(2), 211–222. https://doi.org/10.1037/0022-3514.78.2.211

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*(2), 332–346. https://doi.org/10.1037/0022-3514.75.2.332

Gino, F., Moore, D. A., & Bazerman, M. H. (2008). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School Working Paper, 08-080*.

Gino, F., Moore, D. A., & Bazerman, M. H. (2010). See no evil: When we overlook other people's unethical behavior. In R. M. Kramer, A. E. Tenbrunsel, & M. H. Bazerman

(Eds.), *Social decision making: Social dilemmas, social values, and ethical judgments* (pp. 241–263). Routledge/Taylor & Francis Group.

Giurge, L. M., & Bohns, V. K. (2021). You don't need to answer right away! Receivers overestimate how quickly senders expect responses to non-urgent work emails. *Organizational Behavior and Human Decisions Processes, 167*, 114–128. https://doi.org/10.1016/j.obhdp.2021.08.002

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366–385. https://doi.org/10.1037/a0021847

Gray, K., MacCormack, J. K., Henry, T., Banks, E., Schein, C., Armstrong-Carter, E., Abrams, S., & Muscatell, K. A. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology, 123*(6), 1199–1222. https://doi.org/10.1037/pspa0000310

Gray, K., Ward, A. F., & Norton, M. I. (2014). Paying it forward: Generalized reciprocity and the limits of generosity. *Journal of Experimental Psychology: General, 143*(1), 247–254. https://doi.org/10.1037/a0031047

Hafer, C. L., & Bègue, L. (2005). Experimental Research on Just-World Theory: Problems, Developments, and Future Challenges. *Psychological Bulletin, 131*(1), 128–167. https://doi.org/10.1037/0033-2909.131.1.128

Hatmaker, T., & Lunden, I. (2020, June). Alexis Ohanian steps down from Reddit board, asks for his seat to go to a black board member. *Tech Crunch.* https://techcrunch.com/2020/06/05/alexis-ohanian-steps-down-reddit-board/

Hayes, A. F. (2017). *Introduction to mediation, moderation, and condition process analysis: A regression based approach* (2nd ed.). The Guilford Press.

Haynes, G., & Gilovich, T. (2010). "The ball don't lie": How inequity aversion can undermine performance. *Journal of Experimental Social Psychology, 46*(6), 1148–1150. https://doi.org/10.1016/j.jesp.2010.05.008

Hebl, M. R., Foster, J. B., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin, 28*(6), 815–825. https://doi.org/10.1177/0146167202289010

Jamieson, J. P., Koslov, K., Nock, M. K., & Mendes, W. B. (2013). Experiencing discrimination increases risk taking. *Psychological Science, 24(*2), 131–139. https://doi.org/10.1177/0956797612448194

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

Jones, T. M., (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *The Academy of Management Review, 16*(2), 366–395. https://doi.org/10.2307/258867

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

Jun, S., Chow, R. M., van der Veen, A. M., & Bleich, E. (2022). Chronic frames of social inequality: How mainstream media frame race, gender, and wealth inequality. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(23), e2110712119. https://doi.org/10.1073/pnas.2110712119

Kennedy, J. A., & Schweitzer, M. E. (2018). Building trust by tearing others down: When accusing other of unethical behavior engenders trust. *Organizational Behavior and Human Decision Processes, 149*, 111–128. https://doi.org/10.1016/j.obhdp.2018.10.001

Kish-Gephart, J. J., Detert, J. R., Treviño, L. K., Baker, V. L., & Martin, S. (2014). Situational moral disengagement: Can the effects of self-interest be mitigated? *Journal of Business Ethics, 125*(2), 267–285. https://doi.org/10.1007/s10551-013-1909-6

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

Lönnqvist, J.-E., Irlenbusch, B., & Walkowitz, G. (2014). Moral hypocrisy: Impression management or self-deception? *Journal of Experimental Social Psychology, 55,* 53–62. https://doi.org/10.1016/j.jesp.2014.06.004

Lowery, B. S., Chow, R. M., & Crosby, J. R. (2009). Taking from those that have more and giving to those that have less: How inequity frames affect corrections for inequity. *Journal of Experimental Social Psychology, 45*(2), 375–378. https://doi.org/10.1016/j.jesp.2008.09.010

Lowery, B. S., Chow, R. M., Knowles, E. D., & Unzueta, M. M. (2012). Paying for positive group esteem: How inequity frames affect whites' responses to redistributive policies. *Journal of Personality and Social Psychology, 102*(2), 323–336. https://doi.org/10.1037/a0024598

Matta, F. K., Scott, B. A., Colquitt, J. A., Koopman, J., & Passantino, L. G. (2017). Is consistently unfair better than sporadically fair? An investigation of justice variability

and stress. *Academy of Management Journal, 60*(2), 743–770. https://doi.org/10.5465/amj.2014.0455

Mattila, A. S., Hanks, L., & Zhang, L. (2013). Existential guilt and preferential treatment: The case of an airline upgrade. *Journal of Travel Research, 52*(5), 591–599. https://doi.org/10.1177/0047287513478504

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

Melaku, T. M., Beeman, A., Smith, D. G., & Johnson, B. (2020). Be a better ally. *Harvard Business Review.* https://hbr.org/2020/11/be-a-better-ally

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711–747. https://doi.org/10.1037/0033-2909.130.5.711

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin, 24*(7), 769-783. https://doi.org/10.1177/0146167298247009

Miller, D. T. (2001). Disrespect and the experience of injustice. *Annual Review of Psychology, 52*, 527–553. https://doi.org/10.1146/annurev.psych.52.1.527

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology, 100*(6), 1678–1712. https://doi.org/10.1037/apl0000022

Moberly, N. J., & Watkins, E. R. (2008). Ruminative self-focus and negative affect: An experience sampling study. *Journal of Abnormal Psychology, 117*(2), 314–323. http://dx.doi.org/10.1037/0021-843X.117.2.314

Monin, B., & Merritt, A. (2012). Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 167–184). American Psychological Association. https://doi.org/10.1037/13091-009

Mowday, R.T. (1996). Equity theory: Predictions of behavior in organizations. In R. M. Steers, & L. W. Porter (Eds.), *Motivation & work behavior*, (4th ed., pp. 111–131), McGraw-Hill, New York.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84*, 221–236. https://doi.org/10.1016/S0010-0277(02)00048-3

Norton, M. I., & Sommers, S. R. 2011. Whites see racism as a zero-sum game that they are now losing. *Perspectives on Psychological Science, 6*(3), 215–218. https://doi.org/10.1177/1745691611406922

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology, 61*, 64–78. https://doi.org/10.1016/j.jesp.2015.06.007

Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification* (Vol. 1). Oxford University Press.

Phillips, L. T., & Jun, S. (2022). Why benefiting from discrimination is less recognized as discrimination. *Journal of Personality and Social Psychology, 122*(5), 825–852. https://doi.org/10.1037/pspi0000298

Phillips, L. T., & Lowery, B. S. (2015). The hard-knock life? Whites claim hardships in response to racial inequity. *Journal of Experimental Social Psychology, 61,* 12–18. https://doi.org/10.1016/j.jesp.2015.06.008

Phillips, L. T., & Lowery, B. S. (2018). Herd invisibility: The psychology of racial privilege. *Current Directions in Psychological Science, 27*(3), 156–162. https://doi.org/10.1177/0963721417753600

Phillips, L. T., & Lowery, B. S. (2020). I ain't no fortunate one: On the motivated denial of class privilege. *Journal of Personality and Social Psychology, 119*(6), 1403–1422. https://doi.org/10.1037/pspi0000240

Rattan, A., & Dweck, C. S. (2010). Who confronts prejudice? The role of implicit theories in the motivation to confront prejudice. *Psychological Science, 21*(7), 952–959. https://doi.org/10.1177/0956797610374740

Rattan, A., & Dweck, C. S. (2018). What happens after prejudice is confronted in the workplace? How mindsets affect minorities' and women's outlook on future social relations. *Journal of Applied Psychology, 103*(6), 676–687. https://doi.org/10.1037/apl0000287

Rehg, M. T., Miceli, M. P., Near, J. P., & Van Scotter, J. R. (2008). Antecedents and outcomes of retaliation against whistleblowers: Gender differences and power relationships. *Organization Science, 19*(2), 221–240. https://doi.org/10.1287/orsc.1070.0310

Richman, L., & Leary, M. (2009). Reactions to discrimination, stigmatization, ostracism, and other forms of interpersonal rejection: A multimotive model. *Psychological Review, 116*(2), 365–383. https://doi.org/10.1037/a0015250

Rosette, A. S., & Zhou Koval, C. (2018). Framing advantageous inequity with a focus on others: A catalyst for equity restoration. *Journal of Experimental Social Psychology, 76,* 283–289. https://doi.org/10.1016/j.jesp.2018.03.002

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Schäfer, M., Haun, D. B. M., & Tomasello, M. (2015). Fair is not fair everywhere. *Psychological Science, 26*(8), 1252–1260. https://doi.org/10.1177/0956797615586188

Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Schein, C., Ritter, R. S., & Gray, K. (2016). Harm mediates the disgust-immorality link. *Emotion, 16*(6), 862–876. https://doi.org/10.1037/emo0000167

Sezer, O., Gino, F., & Bazerman, M. H. (2015). Ethical blind spots: Explaining unintentional unethical behavior. *Current Opinion in Psychology, 6*, 77-81. https://doi.org/10.1016/j.copsyc.2015.03.030

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, *24*(2), 125–130. https://doi.org/10.1177/0963721414553264

Shore, T. H. (2004). Equity sensitivity theory: Do we all want more than we deserve? *Journal of Managerial Psychology, 19*(7), 722–728. https://doi.org/10.1108/02683940410559400

Sirola, N., & Pitesa, M. (2017). Economic downturns undermine workplace helping by promoting a zero-sum construal of success. *Academy of Management Journal*, *60*(4), 1339–1359. https://doi.org/10.5465/amj.2015.0804

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

Sull, D., Sull C., Cipolli, W., & Brighenti, C. (2022). Why every leader needs to worry about toxic culture. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/why-every-leader-needs-to-worry-about-toxic-culture/

Swim, J. K., & Hyers, L. L. (1999). Excuse me—What did you just say?!: Women's public and private responses to sexist remarks. *Journal of Experimental Social Psychology, 35*(1), 68–88. https://doi.org/10.1006/jesp.1998.1370

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19*(4), 339–347. https://doi.org/10.1111/j.1467-9280.2008.02091.x

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67–85. https://doi.org/10.1037/0033-2909.110.1.67

Van de Vyver, J., Houston, D. M., Abrams, D., & Vasiljevic, M. (2016). Boosting belligerence: How the July 7, 2005, London bombings affected liberals' moral foundations and

prejudice. *Psychological Science*, *27*(2), 169–177. https://doi.org/10.1177/0956797615615584

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188.

Walster, E., Walster, G. W., & Berscheid, E. (1978). *Equity: Theory and research*. Boston: Allyn and Bacon, Inc.

West, T. V., Magee, J. C., Gordon, S. H., & Gullett, L. (2014). A little similarity goes a long way: The effects of peripheral but self-revealing similarities on improving and sustaining interracial relationships. *Journal of Personality and Social Psychology, 107*(1), 81–100. https://doi.org/10.1037/a0036556

Wilkins, C. L., Wellman, J. D., Babbitt, L. G., Toosi, N. R., & Schad, K. D. (2015). You can win but I can't lose: Bias against high-status groups increases their zero-sum beliefs about discrimination. *Journal of Experimental Social Psychology, 57*, 1–14. https://doi.org/10.1016/j.jesp.2014.10.008

Zitek, E. M., & Jordan, A. H. (2021). Individuals higher in psychological entitlement respond to bad luck with anger. *Personality and Individual Differences, 168,* 110306. https://doi.org/10.1016/j.paid.2020.110306

Zitek, E. M., Jordan, A. H., Monin, B., & Leach, F. R. (2010). Victim entitlement to behave selfishly. *Journal of Personality and Social Psychology, 98,* 245–255. https://doi.org/10.1037/a0017168

Zlotnick, R. (2020, July). White celebs who have given up their roles voicing characters of color. *Distractify.* https://www.distractify.com/p/celebs-gave-up-voicing-characters-of-color