



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Experimental Social Psychology

journal homepage: [www.elsevier.com/locate/jesp](http://www.elsevier.com/locate/jesp)

## The effect of moral loss and gain mindset on confronting racism

Hanna Szekeres<sup>a,b,\*</sup>, Eran Halperin<sup>c</sup>, Anna Kende<sup>a</sup>, Tamar Saguy<sup>d</sup><sup>a</sup> ELTE Eötvös Loránd University, Budapest, Hungary<sup>b</sup> Harvard University, Cambridge, MA, USA<sup>c</sup> Hebrew University, Jerusalem, Israel<sup>d</sup> Interdisciplinary Center Herzliya, Israel

## ARTICLE INFO

## Keywords:

Confronting prejudice  
Moral framing  
Moral mindset  
Regulatory focus theory

## ABSTRACT

In the present research, we tested whether the prospect of moral failure or moral gain can motivate (some) people to confront racism. We investigated the influence of moral loss and moral gain mindset on people's tendency to contest racism as a function of their moral commitment to non-prejudice. Drawing on research on regulatory focus, we predicted that a moral loss mindset (vs. control) would increase confronting tendencies among those who are morally committed to non-prejudice (to safeguard their moral self-concept). A moral gain mindset (vs. control) was expected to increase confronting among those who are less committed to non-prejudice (to enhance their moral self-concept). In Experiment 1, participants were presented with racist scenarios. We varied the framing of moral considerations involved (loss vs. gain vs. control) and assessed confronting intentions. In Experiment 2, participants went through a moral mindset intervention. After a few days, using a behavioral paradigm, we tested their actual confronting during a racist situation. We found partial evidence to our predictions. Those highly committed to non-prejudice (as indicated by a measure of moral conviction in Experiment 1, but by a measure of moral identity in Experiment 2) were significantly more prone to confront in the loss mindset condition than in the control. Confronting in the gain condition was not significantly different than in the control condition at any level of moral commitment to non-prejudice. These findings suggest that a moral loss mindset intervention may be effectively used in promoting (some) people's tendency to speak up against racism.

Elin Ersson, a young university student boarded a plane heading from Sweden to Turkey in July 2018 to protest the deportation of an Afghan asylum seeker, who was forced on that flight. She refused to sit down, preventing take off, until the man was removed from the aircraft. Ersson sacrificed a lot. During the protest she faced an angry cabin crew, complaints by other passengers, and potential legal charges. Eventually, Ersson succeeded in her protest and received an ovation from passengers and was widely praised on social media around the world for her intervention, many calling her a hero. The current research addresses the motives and the mechanisms underlying such confronting behavior by third-party individuals. We investigated how a prospective personal moral failure (of not intervening) or moral gain (of intervening) can play a key role in motivating people like Ersson to perform courageous acts.

Moral courage is a willingness to take a stand in defense of one's own moral principles even when others do not (Miller, 2000; Skitka, 2012). Ersson's action is unique to the extent that the majority of people think they would speak up against intergroup bias, however, often they

do not. For example, even though White Americans anticipated feeling upset and taking action against someone who espouses racial bias against a Black person, those put in that actual situation reported little negative affect and forewent punishing the racist person (Karmali, Kawakami, & Page-Gould, 2017; Kawakami, Dunn, Karmali, & Dovidio, 2009). Similarly, heterosexual participants who imagined witnessing a homophobic slur reported higher intentions of confronting than people who actually witnessed the slur (Crosby & Wilson, 2015). Such inaction is unfortunate because confronting, especially if done by third-party individuals, can actually change perpetrators' beliefs and reduce prejudice (Czopp & Monteith, 2003; Czopp, Monteith, & Mark, 2006).

Research suggest that people fail to exhibit moral courage, and confront intergroup bias because they rather avoid interpersonal costs, such as being perceived as a troublemaker, being disliked or experiencing retaliation (e.g., Eliezer & Major, 2012; Kaiser & Miller, 2001, 2003; Shelton & Stewart, 2004; Swim & Hyers, 1999). While it has been seldom discussed in the confronting literature, omission of intervening can entail personal, psychological costs as well, in the form of deflating

\* Corresponding author at: Department of Psychology, ELTE Eötvös Loránd University, 1064 Budapest, Izabella street 46, Hungary.

E-mail address: [szekeres.hanna@ppk.elte.hu](mailto:szekeres.hanna@ppk.elte.hu) (H. Szekeres).

<https://doi.org/10.1016/j.jesp.2019.103833>

Received 1 October 2018; Received in revised form 25 June 2019; Accepted 26 June 2019

Available online 18 July 2019

0022-1031/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

one's moral self-concept. Indeed, maintaining a non-prejudiced self-image is important to many individuals (Dutton & Lennox, 1974; Monteith, 1993; Plant & Butz, 2006; Plant & Devine, 1998). When individuals' non-prejudiced identity is threatened, they employ different strategies to reinstate it, such as engaging in downward social comparisons with bigots (O'Brien et al., 2010; Wills, 1981), inhibiting prejudiced responses to jokes (Monteith, 1993), or being more generous to an outgroup member (Dutton & Lake, 1973). Thus, we can expect individuals who identify as non-prejudiced to be motivated to confront racism when feeling their moral identity is at risk - which likely is the case when they witness racism and have an opportunity to intervene (i.e., confront). Along the same lines, confronting can also provide personal gains by enhancing one's moral self-concept. Accordingly, individuals might be motivated to confront racism because they construe it as an opportunity for self-enhancement (i.e., desire to increase positive self-concept; Leary, 2007) or self-improvement (i.e., the desire to improve aspects of one's self; Sedikides, 1993; Sedikides & Hepper, 2009).

Thus, we here propose that considerations about one's own morality likely weigh in when deciding whether to confront prejudice or not. Such considerations can be induced in different ways, even by simply priming aspects of moral courage. For example, people associate different moral behaviors with different moral prototypes (helping with being caring, moral courage with being just, heroism with being brave). Accordingly, an activation of a certain prototype (e.g., "just", which is associated with being fair, moral, truthful, honest) was shown to increase the tendency for morally courageous behavior (Osswald, Greitemeyer, Fischer, & Frey, 2010). Extending previous research, we go beyond the moral priming effect to investigate the unique role that moral loss and moral gain mindsets potentially plays in motivating moral behavior, like confronting. According to our thinking, a person with a loss mindset is likely to feel that by not confronting bias, he/she can lose a sense of moral integrity. A person with a gain mindset is likely to feel that by confronting, he/she will earn a sense of being a more moral human being.

While both a moral loss and moral gain mindsets are related to one's moral self-concept, they are also psychologically different and accordingly may motivate confronting in different ways and for different people. Corresponding to these two mindsets, self-regulatory focus theory (Crowe & Higgins, 1997; Higgins, 1997) distinguishes between two motivational systems that regulate goal-directed behavior: a *promotion* and a *prevention focus*. Promotion focus emphasizes advancement and growth, with goals being viewed as ideals (Shah & Higgins, 1997). Prevention focus emphasizes safety, duties and responsibilities, with goals being viewed as obligations. Those with promotion focus are primarily concerned with the presence or absence of positive outcomes (or end states), while those with prevention focus are concerned with negative outcomes. Thus, promotion focus orients people toward pursuing opportunities, whereas prevention focus orients toward avoiding errors (Crowe & Higgins, 1997; Higgins, Roney, Crowe, & Hymes, 1994).

Messages that promote a moral gain mindset regarding confronting (e.g., "intervening would reveal a good and moral side of me") correspond to a promotion focus to the extent that it emphasizes the opportunity of a positive moral end state if the person confronts racism. Similarly, a moral loss mindset (e.g., "not intervening would reveal a bad and immoral side of me") corresponds to prevention focus to the extent it orients people to avoid making an error and ending up with a negative moral state if one fails to confront racism.

Research on regulatory focus suggests that under a prevention focus people are likely to react more strongly to issues related to justice and morality (Sassenberg & Hansen, 2007), especially if they are morally committed to that particular goal. For example, when individuals were primed with a prevention focus (wrote about what they felt they ought to achieve in their working life), the more they held a moral conviction about the fair treatment of their group, the more they supported

collective action against ingroup discrimination (Zaal, Laar, Ståhl, Ellemers, & Derks, 2011). Researchers argued that because prevention-orientation makes people construe goals such as those mandated by moral conviction (in this case, fair treatment) as necessities (Scholer, Zou, Fujita, Stroessner, & Higgins, 2010; Shah & Higgins, 1997; Zaal, Van Laar, Ståhl, Ellemers, & Derks, 2012), they were, presumably, particularly sensitive to the possible losses of inaction and were motivated to avoid those. Meanwhile, for individuals primed with a promotion focus (who wrote about what they would ideally like to achieve) moral conviction did not predict collective action intentions (Zaal et al., 2011). The researchers assumed that promotion-oriented individuals, for whom expectations for success play a key role in taking action (Shah & Higgins, 1997; Zaal et al., 2012) were likely doubtful regarding the effectiveness of collective action. Thus, this work exemplifies how a loss mindset and a gain mindset trigger a different set of concerns, resulting in different actions.

Another study exploring the relationship between moral commitment, regulatory focus and moral behavior found similar results (Brebels, De Cremer, Van Dijke, & Van Hiel, 2011). Business students were made to imagine being managers of a company, and their procedural justice intentions were assessed. Results revealed that participants for whom morality was a central part of their identity, exhibited more procedural justice intentions under a prevention focus (manipulated via priming a threat to the company's position in the market) than under a promotion focus (manipulated via priming an opportunity to advance the company's position; Brebels et al., 2011). One possible explanation is that under a prevention focus, participants focused on the possibility of feeling immoral (loss to moral identity), which they were motivated to avoid. Those for whom moral identity was less important, showed the opposite pattern, i.e. more justice intention in promotion than in prevention focus (Brebels et al., 2011). Perhaps under a promotion focus they were made to think about how they could potentially improve their moral identity by acting fairer.

Together, the work described above suggests that a loss mindset is likely to promote intentions for moral behavior, particularly for those who care about being moral. Applied to our context, this suggests that a loss mindset can promote confronting among those highly committed to non-prejudice, while potential gains might not. This notion echoes prospect theory (Tversky & Kahneman, 1991, 1992), according to which losses inflict psychological harm to a greater degree than gains gratify, which means that people are more willing to run risks to avoid losses than to approach gains. Thus, the psychological costs of falling short of one's moral self-concept should be a motivating force in confronting racism for those who care about being non-prejudiced.

Nevertheless, a loss mindset is not likely to cause change in confronting rate among those weakly committed to non-prejudice, because they should perceive little threat to their non-prejudiced (moral) self-concept as a result of not contesting racism. On the contrary, they might even appraise a loss message as external pressure and obligation to respond without prejudice and thus reduce their intention to confront as a result of a backfire effect (Does, Derks, Ellemers, & Scheepers, 2012; Legault, Gutsell, & Inzlicht, 2011; Powell, Branscombe, & Schmitt, 2005).

Meanwhile, a focus on gains to one's moral self-concept could drive more confronting among those weakly committed to non-prejudice because it is seen as an opportunity to improve moral self-regard (Leary, 2007; Sedikides, 1993; Sedikides & Hepper, 2009). Such opportunity for moral self-improvement should play less of a role in motivating confronting among those who already view themselves as non-prejudiced. Additionally, under moral gain focus, those weakly committed to non-prejudice may confront to gain moral credits prospectively in the domain of racism (Cascio & Plant, 2015) - indeed, prejudiced individuals show higher tendency than non-prejudiced individuals to license their biased/immoral behavior with prior unbiased/moral behavior (Effron, Cameron, & Monin, 2009).

## 1. The present research

Taken together, we predicted that participants' moral commitment to non-prejudice would moderate the effects of moral mindset on confronting racism. Specifically, a moral loss (vs. control) mindset would significantly increase confronting tendencies among those strongly morally committed to non-prejudice, but not among those weakly committed (H1). We also predicted that a moral gain (vs. control) mindset would drive confronting among those who are weakly committed to non-prejudice and would not affect those strongly committed (H2).

We tested these hypotheses in two experiments. In the first, participants were asked to picture themselves in specific racist scenarios, for which confronting was framed in one of three ways: as moral loss, as moral gain, or neither. Participants were asked to report their willingness to confront. In the second study, participants were again randomly assigned to either a loss mindset, a gain mindset, or an empty control using a different manipulation that we considered to be possibly more enduring. After a few days they went through a behavioral paradigm where they witnessed racism and had an opportunity to confront. By investigating whether moral mindsets increase confronting, the present research allowed us to gain insight into different motivations for confronting racism, which can in turn inform interventions aimed at promoting standing up against racism, as well as against other forms of immoral behavior.

Based on the literature on regulatory focus and moral orientation (Brebels et al., 2011; Zaal et al., 2011) we assessed participants' moral commitment to non-prejudice with the moral identity self-importance scale (Aquino & Reed II, 2002), which comprises two dimensions/subscales (internalized vs. symbolized), and with the moral conviction scale (Skitka & Morgan, 2014). Both were slightly altered to the prejudice domain. For an individual with a strong moral identity, moral strivings are integrated with the self-concept and are central to a person's self-definition (Aquino & Reed II, 2002). The internalized component captures a personal and private aspect of moral self-concept and the symbolized component captures the social and public aspect. Higher scores on moral conviction capture individuals' strong and absolute belief that something is wrong or right (Skitka, 2010).<sup>1</sup> We are unaware of previous work that tested these scales together, especially not in the current context. We treated each of the scales (3 in total) as a potential moderator of the predicted effects, and our research was exploratory as to under what condition and which morality orientation would influence the relationship between loss/gain mindsets and confronting.

## 2. Experiment 1

To provide initial support for our predictions, in this study we manipulated moral mindset and measured the self-reported tendency to confront racism. All participants were provided with two vignettes, each depicting an instance of racism (one against a Spanish-speaking boy and another against a Muslim woman, both placed in the US). They were asked to imagine themselves as taking part in these situations, namely, witnessing racism and having opportunity to confront. The opportunity to confront (or not) was manipulated to involve potential

<sup>1</sup> In addition, we included a measure that intended to capture the strength of one's general moral ought vs. ideal orientation (which reflects the tendency to avoid moral failure vs. approach moral ideals, respectively; Aoki, 2015) in order to demonstrate that the predicted mechanisms are specific to moral stance about prejudice and not to general morality. Given that the outcome measure was specific to intergroup situations involving confronting prejudice, we did not expect general moral orientation to moderate the effects. Results supported this expectation. In both studies, we also included an SDO and IMS-EMS scale for exploratory purposes. See Supplementary material for further details on all of these additional scales.

moral self-concept loss, moral self-concept gain or neither. In this latter control group participants were exposed to the same scenarios, and to the opportunity to confront, but no manipulation of loss or gain was added. This enabled us to test the effect of loss and gain beyond a morality priming effect. Then we asked participants about their willingness to confront in the situations described. As hypothesized moderating variables, participants responded to scales measuring the strength of their moral commitment to non-prejudice.

### 2.1. Method

#### 2.1.1. Participants and procedure

We recruited 480 participants via Amazon's Mechanical Turk (mTurk), who participated in an online study for monetary compensation (\$1.20). Power analysis revealed that to detect an (assumed) small effect size (0.2) to achieve a power of at least 0.80 for a moderation analysis (to calculate, we used F test ANOVA/interaction with 6 groups), the suggested sample size was 400 (G\*Power 3.1; see Faul, Erdfelder, Buchner, & Lang, 2009). Given that some participants might be excluded based on attention checks, our collected sample size was 480. Participants were randomly assigned either to the moral loss, moral gain or control group. Participants filled out the morality scales and completed the vignette scenario measures. The materials were counterbalanced such that the morality scales appeared either prior to, or after, the manipulation. At the end of the study, participants responded to demographic questions (age, gender, education level, conservative-liberal orientation; relative socio-economic situation, see Appendix A for full demographics), were debriefed and thanked for their participation.

Data of participants who failed the attention check (which was placed close to the morality scales, "For this question mark number seven as a response";  $n = 33$ ) were excluded, leaving 447 participants for analyses ( $n = 147$  in loss,  $n = 143$  in gain,  $n = 157$  in control condition; 47.3% female, 52.2% male, 0.5% other,  $M_{\text{age}} = 36.38$  years,  $SD_{\text{age}} = 11.56$ ).

#### 2.1.2. Stimuli and measures

For all measures, participants responded on a 9-point scale ranging from 1 (not at all true of me) to 9 (completely true of me) unless indicated otherwise.

**2.1.2.1. Moral conviction.** We used the 4-item moral conviction scale (Skitka & Morgan, 2014). To reflect conviction about prejudice we gave participants the following instruction stem: "To what extent is your position on standing up against prejudice and discrimination...". Participants then responded to the following four items: (1) "a reflection of your core moral beliefs and convictions?" (2) "connected to your beliefs about fundamental right and wrong?" (3) "based on moral principle?", and (4) "a moral stance?" Mean scores on these items were calculated for each participant composing an internally consistent scale ( $\alpha = 0.94$ ). Higher scores indicated higher levels of moral conviction against prejudice.

**2.1.2.2. Moral-prejudice identity.** We used the moral identity (MID) self-importance scale (Aquino & Reed II, 2002) and adapted it to the current context by revising the scale instruction to refer to non-prejudiced values and by excluding one item that did not fit the current context. For the MID internalization subscale participants responded to 5 items, such as "Being someone who has these views and beliefs [being non-prejudiced] is an important part of who I am" ( $\alpha = 0.83$ ). For the MID symbolization subscale participants responded to 4 items, such as "The fact that I have these views and beliefs [being non-prejudiced] is communicated to others by my membership in certain organizations" ( $\alpha = 0.92$ , see Appendix B for full scales).

**2.1.2.3. Vignette scenarios and confronting intentions.** The scenarios and

measures were developed for the purpose of the current study. All participants were presented with the same two scenarios (the order of the two scenarios was counterbalanced). In scenario A, the participant had to imagine that she or he witnesses a man verbally assaulting a Spanish-speaking teenage boy on the bus and expressing his dislike of immigrants (see Appendix C for full texts). In scenario B, the participant allegedly overheard his or her co-workers making fun of their Muslim female co-worker for her religion.

For each scenario, we mentioned a dilemma and asked participants to imagine the pro's (e.g., "you believe that this specific boy is treated unfairly, you are debating whether to intervene or not") and con's of confronting (e.g., "if you get involved, the man may verbally or even physically attack you"). Until this point all participants read the same scenario. The manipulation of moral loss vs. gain was communicated at the end of this text. Specifically, in the loss condition, additional arguments referring to moral considerations framed around losses were presented (e.g., "You feel it is your moral obligation to intervene. If you don't intervene, you fail your moral duty, and you may later feel like a worse person morally. You feel you can lose a lot if you don't confront."). In the gain condition the additional arguments referred to a moral gain (e.g., "You feel it is your moral aspiration to intervene. If you intervene, you succeed to live up to your moral principles, and you may later feel like a better person morally. You feel you can gain a lot if you confront."). In order to encourage participants to carefully read the scenarios we included an open-ended question under each scenario, which read: "Based on the text, what are the considerations in the decision to intervene?". Responses to this question were not analyzed.

Following each scenario, all participants were asked about their confronting intentions: First, we measured willingness to engage in specific confronting actions with six items rated on a 9-point scale (from 1 = not likely at all to 9 = very much likely), such as "I would confront the man and tell him he is racist." or "I would ask the man to stop assaulting the boy." (scenario A) and "I would tell my supervisor about my co-workers' conversation" or "I would ask my co-workers to stop insulting her" (scenario B; see Appendix C for full measure).<sup>2</sup> Then, for each scenario, we also included an item assessing overall confronting willingness: "Overall, to what extent you would confront in this situation in order to [help the boy/stand up for her?]" on a 9-point scale from 1 = I would not confront at all to 9 = I would totally confront. Given that people may vary in the form of confronting they choose to take, and a participant may prefer one way of confronting very much while not at all another, we extracted the highest score each participant gave across the 6 items for each scenario (reverse coded the non-confronting option items). This way, we captured the greatest tendency to confront, for each participant. The two values (maximum value from each scenario) were then averaged with the overall general confronting scores (2 for each scenario) given by each participant. These four numbers formed an internally consistent 'confronting intentions' measure ( $\alpha = 0.75$ ).

Following the vignette scenarios and intention measure, we asked participants in the loss and gain experimental conditions to indicate "Which of the following is closer to what was suggested in the texts about feelings and morality?" Answer options were either 1 = "After confronting, people may feel better and gain positive moral identity" (indicating moral gain) or 2 = "After not confronting, people may feel worse and lose positive moral identity" (indicating moral loss). We considered this a manipulation check.

<sup>2</sup>As the seventh item on these blocks we put: "other suggestion (not mandatory, if you don't write, just mark 1): \_\_\_\_\_ (text entry)". To avoid anchoring, we did not write an item, which would describe an action in agreement with the racist perpetrator (i.e., insulting the boy or the female co-worker). We included this item to provide an opportunity for participants to express this sentiment if they wished. We did not analyze these responses or considered scores on this item in data analyses.

## 2.2. Results

### 2.2.1. Preliminary analyses

Using One-way ANOVA and post-hoc tests we did not find significant differences between conditions on demographic variables,  $p$ 's > .14 (see Table 1 for means, standard deviations and correlations between study variables). However, there were significant differences between conditions on the moral conviction scale (loss vs. control:  $p = .019$ , gain vs. control:  $p = .004$ ),  $F(2,444) = 4.74$ ,  $p = .009$ . There were no significant differences on the MID-symbolization scale (loss vs. gain:  $p = .075$ , gain vs. control:  $p > .25$ ;  $F(2,444) = 1.66$ ,  $p > .19$ ), nor on the MID internalization scale ( $p$ 's > .10;  $F(2,444) = 1.58$ ,  $p > .20$ ). We also tested whether the order of study materials had an effect on confronting intentions or on the morality scales and found non-significant differences ( $p$ 's < .12). In addition, we tested and found no significant two-way interactions between order and condition (loss vs. gain vs. control) on confronting intentions ( $p > .25$ ) or on the morality scales ( $p = .093$  for MID-symbolization, otherwise  $p$ 's > .25).<sup>3</sup> Given these results we nonetheless decided to control for order as a covariate in our main analyses.

As a next step, we tested the manipulation check item. As intended, we found that in the loss condition, significantly more participants chose the loss response ("... people may feel worse and lose positive moral identity"; 66.2%) over the gain response, and in the gain condition, significantly more participants indicated the gain response over the loss response ("... people may feel better and gain positive moral identity"; 91.5%);  $\chi^2(1) = 98.96$ ,  $p < .001$ , Cramer's  $V = 0.60$ .

### 2.2.2. Main analyses

In order to analyze the effects of a loss mindset (vs. control) and of a gain mindset (vs. control) on confronting intentions as a function of participants' moral commitment to non-prejudice, we ran a moderation analyses for a multicategorical IV (Hayes, 2018). The analysis involved two dummy variables as independent variables: D1 (1 = loss and 0 = control and gain) and D2 (1 = gain and 0 = control and loss). The morality scales, namely, moral conviction, MID-internalization and MID-symbolization were tested as moderators, each in a separate model. Variables were not z-standardized or centered. As indicated before, we controlled for order of study materials (as covariate) in all three moderation analyses.

In the analysis where moral conviction was tested as a moderator, the two-way interaction between D1 (loss vs. control) and moral conviction on confronting intentions was not significant ( $p = .057$ ; see Table 2 for statistics for the interactions as well as the conditional main effects of both dummy variables and morality scales). Simple effects revealed that the loss mindset affected only those who were high on moral conviction (1 SD above the mean), such that it increased confronting intentions ( $M = 7.99$ ) compared to the control condition ( $M = 7.48$ ),  $b = 0.52$ ,  $SE = 0.25$ ,  $t = 2.11$ ,  $p = .035$ , 95% CI [0.04, 1.00] (see Fig. 1). For those weakly morally convicted (1 SD below mean), this effect was not significant ( $p > .25$ ).

The MID symbolization and internalization subscale did not moderate the relationship between loss (vs. control) mindset and confronting ( $p$ 's > .25), and simple effects were also not significant ( $p$ 's > .25). Additionally, there were no significant two-way interactions between D2 (gain vs. control) and any of the moral commitment scales ( $p$ 's > .25; and simple effects were also non-significant,  $p$ 's > .25). (See Table S1 in Supplementary material for estimated conditional means and simple effects for confronting intentions as a factor of condition and all moral commitment to non-prejudice scales).

<sup>3</sup>The three-way interaction between order, condition (loss vs. gain vs. control), moral commitment to non-prejudice scales (analyzed separately) on confronting were not significant either ( $p$ 's > .25).

**Table 1**  
Means (standard deviations) and correlations between study variables in Experiment 1.

	<i>M (SD)</i>	Confronting intentions	Moral conviction	MID-internal	MID-symbol	Conservative-liberal	SES	Education	Age
Confronting intentions	7.02 (1.63)	–							
Moral conviction	7.29 (1.76)	0.43**	–						
MID-internal	7.33 (1.67)	0.26**	0.53**	–					
MID-symbol	5.28 (2.23)	0.29**	0.37**	0.31**	–				
Conservative-liberal	6.29 (2.96)	0.18**	0.21**	0.25**	0.06	–			
SES	3.40 (1.01)	0.11**	0.04	–0.13**	0.18**	–0.14**	–		
Education	2.79 (0.73)	0.11*	–0.005	–0.05	0.12*	0.05	0.30**	–	
Age	36.38 (11.56)	–0.03	0.10*	0.19**	0.01	–0.10*	–0.16**	0.07	–

Confronting intentions, moral conviction, and MID's were on a 9-point continuous scale. Conservative-liberal dimension was on a continuous slider from 0 (conservative) to 10 (liberal). SES was on 6-point and education was on a 5-point ordinal scales.

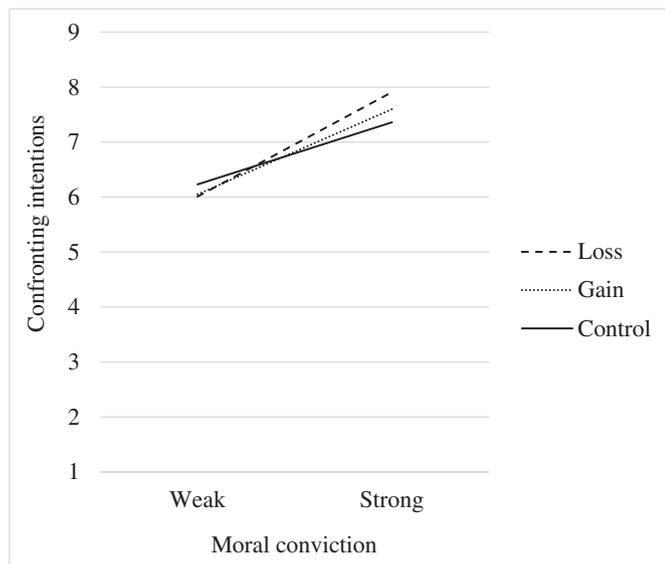
Note.

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table 2**  
The effect of moral mindset condition (control, loss, gain) on confronting intentions as a factor of moral commitment to non-prejudiced scales (controlling for order) in Experiment 1.

Moderator	Predictor	Confronting intentions			
		B (SE)	<i>t</i>	<i>p</i> -Value	95% CI
Moral conviction	Moral conviction	0.30 (0.08)	3.80	.00	0.15; 0.46
	D1 (control vs. loss)	–1.34 (0.82)	–1.63	.10	–2.96; 0.28
	D2 (control vs. gain)	–0.56 (0.77)	–0.73	.47	–2.06; 0.94
	D1 x moral conviction	0.21 (0.11)	1.91	.06	–0.01; 0.42
	D2 x moral conviction	0.09 (0.10)	0.85	.39	–0.11; 0.28
	Order	0.06 (0.16)	0.37	.71	–0.25; 0.37
MID-symbol	MID-symbol	0.24 (0.06)	4.02	.00	0.12; 0.36
	D1 (control vs. loss)	0.41 (0.50)	0.82	.41	–0.57; 1.39
	D2 (control vs. gain)	–0.06 (0.47)	–0.12	.91	–0.97; 0.86
	D1 x MID-symbol	–0.09 (0.09)	–1.04	.30	–0.26; 0.08
	D2 x MID-symbol	–0.01 (0.08)	–0.11	.91	–0.17; 0.15
	Order	0.02 (0.17)	0.10	.92	–0.31; 0.34
MID-internal	MID-internal	0.27 (0.07)	3.77	.00	0.13; 0.42
	D1 (control vs. loss)	0.17 (0.86)	0.20	.84	–1.51; 1.86
	D2 (control vs. gain)	0.42 (0.80)	0.52	.60	–1.16; 2.00
	D1 x MID-internal	–0.02 (0.11)	–0.20	.84	–0.24; 0.20
	D2 x MID-internal	–0.07 (0.11)	–0.62	.54	–0.28; 0.14
	Order	–0.05 (0.17)	–0.29	.77	–0.38; 0.28



**Fig. 1.** Interaction between mindset framing condition (loss vs. gain vs. control) and participants' moral conviction on confronting intentions (on a 9-point scale).

2.3. Discussion

In Experiment 1, we found partial support for the idea that a moral mindset induction can increase levels of confronting racism. Consistent with our prediction (H1), we found that the moral loss framing, compared to a control, triggered more willingness to confront racism among those who were high on moral commitment to non-prejudice, specifically on moral conviction. There was no significant effect among those who were weakly convicted. The interaction, nevertheless, was not significant,  $p = .057$ , and these effects occurred with only one out of three potential moderators (the MID subscales did not have an influence on the relationship between loss mindset and confronting).

Furthermore, we failed to find evidence for our second prediction (H2) regarding gain mindset. We found that confronting rate in the moral gain condition did not significantly differ from the control condition, at any level of moral conviction or MID-symbolization or MID-internalization. It could be the case that the gain manipulation was not effective because, overall, our moral mindset priming was subtle and perhaps not sufficiently persuasive. In the next study, we created and tested an engaging and distinctive moral mindset intervention. We are also not sure whether participants perceived the vignette situations as depicting prejudice given that we did not have a proper measure to assess that. Additionally, we measured (hypothetical) willingness to confront racism and not actual behavior. Once participants need to make an (allegedly) real decision to confront, a different pattern of

results may emerge (Crosby & Wilson, 2015). To overcome these limitations of the confronting measure, in Experiment 2, we employed a behavioral test of confronting racism, where participants believed that they were actually witnessing blatant prejudice and had an opportunity to contest it.

### 3. Experiment 2

Experiment 2 involved an online intervention we designed to induce a moral loss vs. gain moral mindset. We further developed a behavioral paradigm to measure actual confronting behavior. Participants first filled out the scales of moral commitment to non-prejudice (same as those in Experiment 1). Then they were randomly assigned either to a moral loss or moral gain mindset intervention, or to an empty control condition. After a couple of days, we approached the same participants with a study allegedly pre-testing a behavioral economics game (which actually included the confronting measure). Participants believed that they were observing a game involving other participants. During the game, they witnessed a player being prejudiced and discriminating against an outgroup member (a Muslim individual) and had an opportunity to respond and thereby confront the racist player. Testing actual confronting allowed us to potentially capture real-life behavior.

#### 3.1. Method

##### 3.1.1. Participants and procedure

We recruited 450 U.S. residents through mTurk to complete the first part of the study for monetary compensation (\$1.50). We ended up with two additional respondents, which is not unusual with mTurk. A-priori power analysis (G\*Power 3.1; see Faul et al., 2009) for logistic regression (probabilities set to 0.25 and 0.15; R-squared other  $X = 0.50^1$ ) revealed that a sample size of 247 is needed to achieve sufficient power of 0.80. Considering attrition and the exclusion criteria (attention and suspicion check), we collected 450 participants in the first part. In the survey, participants first completed the morality scales, then they were randomly assigned to a moral loss or moral gain mindset induction, or empty control condition (in which they completed the morality scales but were not exposed to any mindset related stimuli). This part of the study ended with demographic questions (*age, gender, SES, education level, liberal-conservative orientations, race/ethnicity and religion*). See Appendix A for full demographics).

Two days later, all respondents received a notification email advertising a new study (allegedly pre-testing a behavioral economics game). This email came from a different mTurk account in order to disguise that the studies were connected.<sup>4</sup> Only two respondents who identified as Muslim in the first part of the study were not invited back to due to ethical considerations (risk of psychological harm). Following our email advertising the new study, we left the study open for 5 days to collect responses. With a 34% dropout rate, 297 respondents returned and went through the confronting measure. Data of participants who failed the attention check in the first part (same 1 item as in Experiment 1,  $n = 27$ , 9.1% of sample) or expressed suspicion about the game stimuli<sup>5</sup> ( $n = 10$ , 3.4% of sample) were excluded from data analyses,

<sup>4</sup> We contacted participants and collected data through an mTurk extension website called TurkPrime (Litman, Robinson, & Abberbock, 2016). We as researchers did not handle participants' identifying information such as their email address, we solely used their anonymized ID's.

<sup>5</sup> We screened for suspicion about the realness of the game based on participants' messages to the alleged player they were observing (whether the participant asked or stated if the player is "real"/"bot") and based on an open-ended question at the end of study ("Please feel free to leave any comment/s or remark/s you may have."; whether participants wrote that the game was fabricated and/or that the racist remark was the actual aim of the study). Note that no participant made a comment that would lead us to believe that they figured out that the two parts were connected.

leaving 260 participants in the study ( $n = 82$  in loss,  $n = 78$  in gain,  $n = 100$  in control condition<sup>6</sup>; 44.6% female,  $M_{age} = 36.08$ ,  $SD_{age} = 10.27$ ).

After completing the study, we messaged all participants for debriefing. We revealed to them the study purpose (how people react to uncomfortable intergroup situations, such as witnessing racism), the deception (that no racism had occurred as the game was pre-programmed), we reassured them the study was anonymous, and we provided them with our email address for further assistance.

##### 3.1.2. Materials and measures

We included the same morality scales, moral conviction (4 items,  $\alpha = 0.94$ ), MID internalization (5 items,  $\alpha = 0.84$ ) and MID symbolization scales (4 items,  $\alpha = 0.89$ ) as in Experiment 1. The rest of the materials are described below.

**3.1.2.1. Intervention stimuli.** In the loss and gain conditions, participants were told that they would be asked to complete three tasks (see Appendix D for full material). Each of the tasks was aimed to induce a loss or gain mindset with respect to failure to confront immoral behavior (one task was general, the other two were specific to racism). In the first task a poster appeared depicting a bystander situation (someone being physically attacked while others around do nothing) with a text either framed as moral loss ("Not getting involved sometimes means you are risking to behave immorally") vs. gain ("Not getting involved sometimes means you are missing a chance to behave morally"), based on condition. Participants were asked to write what they thought the poster meant. In the next task, participants were shown a video of a real event, depicting a British woman insulting immigrants on a bus, and then a text described a bystander passenger, who later allegedly reported his regrets of not confronting. Participants were asked to give a short account of their thoughts and feelings while imagining they are this passenger. The provided text box started with a stem sentence, which was framed according to condition (moral loss: "I feel like not intervening revealed a bad side of me..." vs. moral gain: "I feel like intervening would have revealed a good side of me..."). In the last task, a text described a Holocaust rescuer and an alleged account he gave about his actions, which was framed according to condition (moral loss: "...He once noted that not doing what he did would have cost him his moral virtue and he would have felt like a bad person..." vs. moral gain: "...He once noted that through this action he gained moral virtue and he feels he became a better person for doing it..."). Then participants were asked to describe this person's potential thoughts and feelings about his own behavior.

Similar to Experiment 1, for a manipulation check, on a separate page we asked participants in the two experimental conditions the following questions: "What is suggested in the previous tasks about people's feeling and morality if they do not intervene in those situations? (pick the convenient sentence starter and continue the sentence)". Answer options were 1 = "they miss a chance to gain... (text entry)", or 2 = "They risk to lose... (text entry)". The chosen option enabled us to assess whether we succeeded to prime a gain or loss mindset (we did not analyze how they completed the sentence).

**3.1.2.2. Confronting stimuli.** We developed the current "Trust Game" paradigm for the purpose of our research program on confronting racism. We used and altered the "Trust Game" (Berg, Dickhaut, & McCabe, 1995; Charness & Rabin, 2002), where Player A decides how much money out of an initial endowment to send to another subject, Player B. The sent amount is then multiplied by 3 and Player B decides

<sup>6</sup> There are more participants in the empty control condition than in the loss and gain conditions most likely because more respondents started these conditions without finishing it than respondents in the control (which was a shorter survey), but Qualtrics still calculated it toward equal randomization.

how much of the money received to send back to Player A. To conceal the purpose of the study, participants were told that we are testing how observing influences trusting behavior and how does gender (of the players and observers) influence trusting behavior. Participants were given instructions on the game (see Appendix E) and they were “trained” on the rules. They were told that they would first be assigned to a player and observe his/her rounds and only after they would play this game themselves for money. Participants were further explained that this player they observe could initiate private messaging with them. Allegedly this was enabled in order for the “observer” to feel more real to the player, in reality this was done in order to manipulate racism and enable confronting. In order to strengthen our cover story, we asked participants not to share the purpose of the study with the players while they were messaging.

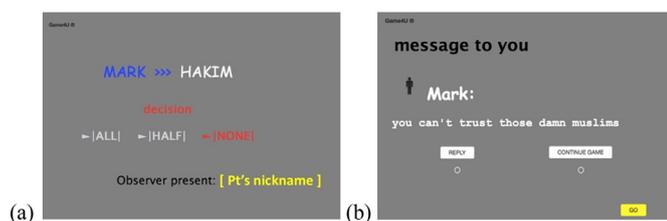
Participants then entered a different site to observe the game (in reality the game observed was pre-programmed). To make participants feel present in the situation we asked them to provide their nickname as well (they appeared with this throughout the game). All participants were assigned to observe a player called Mark, and then observed two decoy rounds and exchanged some messages with Mark (in these messages Mark addressed the participant by their nickname and his responses were written to fit any message the participant replied with). In the first observed round Mark played against Kip and gave half of his money, then against Nica and gave all of his money (both partners returned the money fairly). Then, Mark's partner appeared as Hakim (a Muslim name), and to him Mark decided to give no money. Then, Mark privately messaged the participant saying: “You can't trust those damn Muslims” (see Fig. 2a and b for scenes from the game). Participants thus witnessed a discriminatory act and an explicitly racist comment. Beneath the message, participants had a chance to either press ‘continue game’ or ‘reply’. Then, for all participants a message appeared on the screen indicating that there was a problem registered in the system, and the game terminated.

**3.1.2.3. Confronting.** Participants who chose to continue the game, or to reply but left the message box empty, or those who responded in a non-confronting way - were all labeled as ‘not confronting’ and coded as ‘0’. Responses that questioned or reproached the player for his behavior and statement were labeled as ‘confronting’ and coded as ‘1’. Responses that are unclear as to their intentions (confronting or not) were coded as other and treated as missing values in the main analyses (as indicated in the participants section part, those who communicated suspicion here about the study were excluded from analyses). These responses were coded by two authors of the manuscript blind to conditions, and disagreements ( $n = 3$ ) were discussed.

## 3.2. Results

### 3.2.1. Preliminary analysis

First, we tested and found no significant differences across conditions on the morality scales,  $p$ 's  $> .17$ , or on demographics,  $p$ 's  $> .22$  (see Table 3 for means, standard deviations and correlations between



**Fig. 2.** Scenes from the game: (a) the observed player/Mark was playing against a Muslim player (Hakim) and decided to give him no money. The observer/participant was indicated to be present, and (b) the observed player/Mark messaged the participant with his prejudiced remark about Muslims.

study variables). We found that the mindset manipulation was successful in communicating the sense of moral loss vs. gain,  $\chi^2(1) = 37.95$ ,  $p < .001$ , Cramer's  $V = 0.49$ . Specifically, significantly more people indicated the loss (vs. gain) response in the loss condition (76.8%), and significantly more participants indicated the gain (vs. loss) response in the gain condition (71.8%).

Next, we analyzed the responses of the confronting message. Considering all conditions, we found that 120 participants pressed to continue the game (were thus coded as not confronting, ‘0’), 140 pressed to reply. Among repliers, 108 participants (41.5% of all participants) confronted the racist perpetrator (coded as confronting, ‘1’), 19 people wrote messages that expressed consent or simple acknowledgment (e.g., “OK”) (they were all coded as not confronting, ‘0’), and 13 responses were ambiguous and thus coded as ‘other’ and were not used for data analyses.

### 3.2.2. Main analyses

Analysis strategy was the same as in Experiment 1, except this time we ran logistic regression because our DV was dichotomous (confronting or not). As indicated in Table 4, there was a significant two-way interaction between D1 (loss vs. control) and MID-symbolization on confronting intentions. A simple-effects analysis revealed that the loss mindset affected only those who were high on MID-symbolization (1 SD above the mean), such that it increased confronting intentions (prob. = 0.56, odds = 1.27) compared to the control condition (prob. = 0.30, odds = 0.43),  $b = 1.10$ ,  $SE = 0.46$ ,  $Z = 2.37$ ,  $OR = 0.34$ ,  $p = .018$ , 95% CI [0.19; 2.01] (see Fig. 3). This effect was not significant among those weakly committed (1 SD below mean;  $p > .25$ ).

Other morality scales (MID internalization and moral conviction) did not moderate the relationship between loss (vs. control) and confronting ( $p$ 's  $> .25$ ), and simple effects were also not significant ( $p$ 's  $> .24$ ). Additionally, there were no significant two-way interactions between D2 (gain vs. control) and any of the moral commitment scales ( $p$ 's  $> .10$ ), nor significant simple effects ( $p$ 's  $> .14$ ). (See Table S2 in Supplementary material for probabilities and simple effects for confronting as a factor of condition and moral commitment to non-prejudice scales.)

## 3.3. Discussion

In Experiment 2, we developed and tested an online moral mindset intervention and a couple of days later employed an online behavioral paradigm to measure real acts of confronting racism. We found partial support for our prediction (H1) that a moral loss mindset can increase confronting of racism among those high in moral commitment to non-prejudice. Nevertheless, also in Experiment 2 the effect emerged only with one out of the three considered moderators, and, unlike in Experiment 1, in this study a MID subscale (symbolization) moderated the effect of moral loss mindset on confronting (and not moral conviction). We discuss potential reasons for this inconsistency in detail in the general discussion. As in Experiment 1, confronting rate was not significantly affected by the gain mindset (vs. control) at any level of moral commitment. We address the lack of support for our expectation in this respect (H2) in the general discussion.

Interestingly, MID-symbolization seemed to have a negative relationship with confronting. Specifically, in the control group, confronting tended to be higher among those low in MID-symbolization compared to those higher on this scale. Note that this scale asked about the respondents' behavioral commitment to non-prejudice in everyday life, about their hobbies, activities, memberships in organizations that reflect these values. Those high on symbolization are usually driven to publicly exhibit their moral self, and they are motivated by recognition and reputational gains from engagement in moral deeds (Winterich, Aquino, Mittal, & Swartz, 2013; Winterich, Mittal, & Aquino, 2013). It could be the case that participants high on this measure felt excused from confronting because their public activities provided them with

**Table 3**  
Percentages or means and standard deviations (M and SD) and correlations between study variables in Experiment 2.

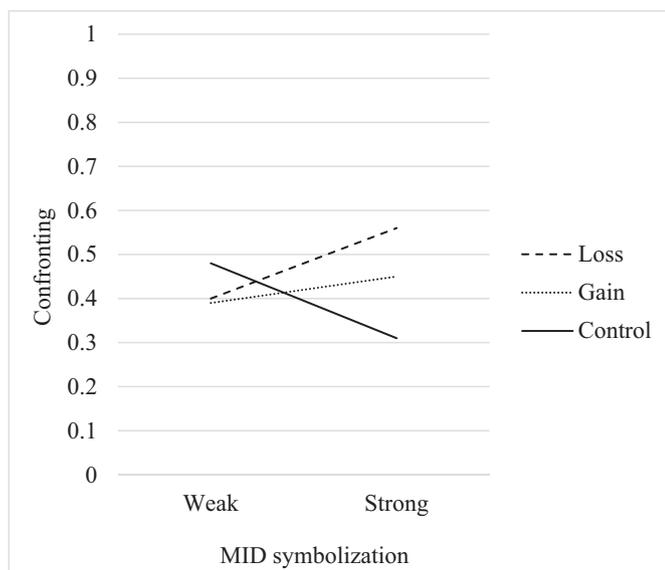
	M (SD)	Confronting (0: not, 1: yes)	Moral conviction	MID-internal	MID-symbol	Conservative-liberal	SES	Education	Age
Confronting (0: not, 1: yes)	56.3%, 43.7%	-							
Moral conviction	7.39 (1.67)	0.16*	-						
MID-internal	7.33 (1.68)	0.21**	0.52**	-					
MID-symbol	5.87 (1.97)	-0.03	0.31**	0.19**	-				
Conservative-liberal	6.38 (3.11)	0.11	0.24**	0.14**	-0.01	-			
SES	3.45 (0.96)	-0.12	-0.02	-0.25**	0.19**	-0.10	-		
Education	2.81 (0.70)	-0.06	-0.07	-0.18**	0.06	-0.05	0.34**	-	
Age	36.08 (10.27)	0.15*	0.11	0.22**	-0.01	-0.08	-0.07	0.10	-

Moral conviction and MID scales were on a 9-point continuous scale. Conservative-liberal dimension was on a continuous slider from 0 (conservative) to 10 (liberal). SES was on 6-point and education was on a 5-point ordinal scales.

Note.  
\*  $p < .05$ .  
\*\*  $p < .01$ .

**Table 4**  
The effect of moral mindset condition (loss, gain, control) on confronting behavior (0: not confronting, 1: confronting) as a factor of the moral commitment to non-prejudiced scales in Experiment 2.

Moderator	Predictor	Confronting behavior			
		B (SE)	Z	p-Value	95% CI
Moral conviction	Moral conviction	0.21 (0.15)	1.36	0.18	-0.09; 0.51
	D1 (control vs. loss)	-0.38 (1.66)	-0.23	0.82	-3.63; 2.88
	D2 (control vs. gain)	0.66 (1.59)	0.42	0.68	-2.44; 3.77
	D1 x moral conviction	0.10 (0.22)	0.47	0.64	-0.32; 0.53
	D2 x moral conviction	-0.08 (0.21)	-0.37	0.71	-0.48; 0.33
MID-symbol	MID-symbol	-0.20 (0.10)	-2.01	0.05	-0.39; -0.01
	D1 (control vs. loss)	-1.76 (1.02)	-1.72	0.09	-3.76; 0.25
	D2 (control vs. gain)	-1.43 (1.01)	-1.41	0.16	-3.41; 0.56
	D1 x MID-symbol	0.37 (0.17)	2.18	0.03	0.04; 0.70
	D2 x MID-symbol	0.27 (0.16)	1.65	0.10	-0.05; 0.59
MID-internal	MID-internal	0.23 (0.14)	1.65	0.10	-0.04; 0.51
	D1 (control vs. loss)	0.02 (1.58)	0.01	0.99	-3.08; 3.12
	D2 (control vs. gain)	-0.75 (1.64)	-0.45	0.65	-3.97; 2.47
	D1 x MID-internal	0.05 (0.21)	0.23	0.82	-0.36; 0.45
	D2 x MID-internal	0.11 (0.21)	0.54	0.59	-0.30; 0.53



**Fig. 3.** Interaction between mindset framing condition (loss vs. gain vs. control) and participants' MID-symbolization on confronting (yes or no; visualizing probabilities).

moral credentials (see Monin & Miller, 2001). This might have not been the case for those who had no such credentials (i.e., low on MID-symbolization). Thus, what the loss mindset intervention possibly did, is

that it induced individuals high on MID-symbolization to think about losing these moral credits, thereby motivating their confronting. Put differently, it is possible that the moral loss mindset acted as a buffer to this general moral licensing process.

Finally, we should point out that our manipulation check, assessing the extent to which we succeeded to communicate a loss/gain mindset was limited because it did not involve a comparison with the control condition. In the manipulation check we asked: "What is suggested in the previous tasks about people's feeling and morality if they do not intervene in those situations?". Answer options were either "they miss a chance to gain... (text entry)", or "they risk to lose... (text entry)". This manipulation check question would have not made sense to participants in the empty control condition because they were not exposed to any tasks.

#### 4. General discussion

People who intervene in times of racial and ethnic discrimination often describe their actions as driven by a need to avoid a sense of moral failure. Holocaust rescuers often explained their decision to help along the line of "Can I live with myself if I say no?" (Fogelman, n.d.). Similarly, in the example of protest against deportation, Ersson herself said in an interview, "I knew that I couldn't back down because it was my name that was on the ticket. I had to do what I could". In the present research, we found evidence that this sense of moral failure can indeed motivate people to confront racism. Across two studies, we tested the effects of thinking about moral loss and moral gain on contesting racism in light of people's non-prejudice moral commitment. In Experiment 1,

participants were presented with vignettes depicting racist scenarios, where we varied the description of potential moral concerns (loss vs. gain vs. control) and assessed participants' self-reported intentions to confront the racist act. In Experiment 2, participants went through a moral mindset intervention (that was intended to induce a loss or gain mindset), or empty control, and a few days later, we employed a behavioral paradigm to measure real action to confront racism.

The studies provide partial support to our predictions. First, we failed to find support for the hypothesis regarding the moral gain mindset (H2). This mindset did not seem to be effective in increasing confronting rate in comparison to the control group at any level of moral commitment to non-prejudice. Regarding our other hypothesis (H1), we predicted and found partial evidence that moral framing can affect the tendency to confront racism, and this is dependent on participants' non-prejudiced moral commitment. Across studies, among those with high moral commitment to non-prejudice, a loss mindset led to more confronting, compared to the control condition (H1). Likely, the loss framing activated motivation to safeguard one's moral non-prejudiced self-concept (Dutton & Lennox, 1974; Monteith, 1993). However, in each of the studies a different variable moderated the effect. In Experiment 1, the moral conviction about prejudice scale (adapted from Skitka & Morgan, 2014) moderated the effect of loss mindset (vs. control) on confronting (although the interaction was not significant, only the relevant simple effect), while in Experiment 2 it was the symbolization facet of the moral identity-prejudice scale (adapted from Aquino & Reed II, 2002) that significantly moderated these effects. We employed these scales and we did not have specific prediction as to (when and) which moral commitment construct would have influence on the relationship between moral gain/loss mindsets and confronting.

Due to these different moderation effects, our study findings may be incidental (and reflect false positive results), and thus attempts of replication are advisable in the future. However, there were notable differences between the two studies that may account for the differing effects of the morality scales. Namely, the manipulation in Experiment 1 involved an imagined scenario that focused on the victim of prejudice, and the participant was not the actual person in the described situation who had the responsibility to confront. In Experiment 2, the prejudiced situation was perceived as real (and not hypothetical), making the participants believe they had to make an actual choice to confront or not, rendering less focus on the harm done to the victims and more focus on the responsibility and actions of the participant. Correspondingly, those high on MID-symbolization are usually driven to publicly exhibit their moral self, and they are motivated by recognition and reputational gains from engagement in moral behavior (Schaumberg & Wiltermuth, 2014; Winterich, Aquino, et al., 2013; Winterich, Mittal, & Aquino, 2013). Thus, in Experiment 2, it was those high on MID-symbolization who became encouraged to confront, perhaps because they felt personally involved in the (allegedly) real-life situation and under loss induction they felt that their own moral public identity and reputation is at risk. Even more so if they thought that others may see their actions such as the perpetrator, other players, or the experimenter.

In the same time, compared to MID-symbolization, moral conviction is a relatively other-oriented moral attitude given that it reflects internally entrenched beliefs (Skitka, 2010; Skitka, 2014; Skitka, Bauman, & Sargis, 2005), which likely renders more focus on the actual subject of this strongly held belief, in our case on the target of prejudice. Accordingly, we speculate that in Experiment 1 where there was a stronger focus on the victims of prejudice, the moral loss induction could trigger specifically those high on moral conviction to confront. Having said that, to our knowledge no prior research have tested or discussed these moral constructs in the same work, therefore our outlined theoretical distinction was speculative. The current explanations to why different morality constructs pertained more to confronting under differing situational cues warrants further investigation as in the

current research we are unable to answer that.

While the empirical evidence to our proposed effect is limited, we offer an important initial step toward investigating the understudied effect of anticipated moral cost on not confronting prejudice. Our findings partially align with research on regulatory focus, which indicates that individuals under prevention focus (corresponding to the loss mindset) are more likely to engage in action aimed at amending injustice directed toward their own group, than those in a control group, and this is not the case for those under promotion focus (corresponding to a gain mindset; e.g., Sassenberg & Hansen, 2007; Zaal et al., 2012). This effect is more pronounced if individuals hold a strong moral conviction about the fair treatment of their group (Zaal et al., 2011). We found consistent pattern in the domain of morality in the context of third-party intervention, showing that those induced to think in terms of a loss to their morality, were more likely to confront that those in a control condition, if they were committed to non-prejudice.

For people with promotion focus, taking action depends heavily on the (perceived) instrumentality of the action, i.e., on the expectation of success (Shah & Higgins, 1997). Therefore, trying to motivate action through reframing the action's moral goal in promotion-oriented terms would only be effective when the likelihood that the action will succeed is high (Quinn & Olson, 2011; Zaal et al., 2012). In our studies, the way we framed the moral gain mindset did suggest that if the individual does confront, he or she would likely succeed in gaining a positive and moral self-regard. At the same time, we did not (necessarily) communicate that confronting will be successful in for example, changing the perpetrator's mind or in helping the victim. This is possibly why we did not find the gain mindset affecting confrontation of racism.

Our findings also reflect, to some extent, the loss aversion effect, which states that losses inflict psychological harm to a greater degree than gains gratify (Tversky & Kahneman, 1991, 1992). However, when people witness racism and contemplate whether to confront, the potential moral self-concept loss may not actually (psychologically) equal the potential gain. Prospect theory was, for the most part, applied to constructs (such as monetary investment) that can be readily quantified. Individuals' relation to their own moral self-concept is not necessarily the same as relations to their material possessions. Thus, the extent to which we can apply loss aversion theory to the current intergroup context is debatable.

Our findings however cannot be explained by a moral priming effect, whereby activating certain aspects of morality in memory (e.g., a just prototype; Osswald et al., 2010) increases morally courageous behavior. For one, in both studies, all participants (including those in the control group) responded to scales that were explicitly about morality. Secondly, in Experiment 1, participants in the control group read the same vignettes that included the prejudicial situation, the opportunity and the pro and con concerns of intervening (while moral gain and loss arguments were not mentioned). Finally, our results showed that the gain mindset manipulation did not influence confronting, and moral commitment moderated only the effects for the loss mindset – rendering it unlikely that priming, or experimental demands can explain our findings.

In general, when it comes to the question of witnessing racism, lay and empirical discussion is usually focused on the personal costs of confronting. Namely, on people's courage to stand up against injustice despite the anticipation of substantial costs to themselves. Such sacrifices are without question admirable and should be recognized. Nevertheless, not much is being said about the personal benefits of confronting, or more correctly, about the personal moral costs of *not* confronting. The present work sheds light on this perspective, by showing that when one cares about being non-prejudiced, the potential loss of one's sense of morality if action is not taken can actually trigger confronting behavior.

Given that confronting in our studies was influenced by the person's consideration about their own morality, and not only about standing up for the victims, can we still consider it a morally courageous behavior?

This resonates with the age old question about the nature and existence of selfless good deeds (Kant, 1785; Nietzsche, 1878), and whether if an individual benefits from their prosocial behavior is that act ultimately egoistic (self-oriented; Andreoni, 1990) or it may nonetheless be considered courageous and altruistic (other-oriented; Batson, 2011; Batson & Shaw, 1991). This remains an unanswered philosophical question. However, considering the motivation of Holocaust rescuers and of Ersson that was mentioned before, we believe that a person's concern about their own morality, triggered by the treatment of *another group* is still at some level an other-oriented concern, that could benefit victims, and mitigate bias among perpetrators. Thus, when considering the tangible outcomes of confronting, we see it as socially beneficial, even if the motive was egoistic.

#### 4.1. Limitations and future directions

Following the previous argument, one could also question whether confrontation in our studies were morally courageous in the sense of involving personal costs to participants. In Experiment 1, in both imagined scenarios we explicitly stated the personal costs involved to confronting (e.g., jeopardize your position and respect at work; be verbally or physically attacked). In Experiment 2, in an (allegedly) real online situation participants were likely concerned that if they confront, they might "lose face", or the perpetrator may reply aggressively, or they sabotage the game and will not get their money. Note, that one study tested a hypothetical situation and the other one was online, thus the generalizability of the findings is limited in this respect. Future research where the study predictions are tested in an in-person, offline context is needed.

Furthermore, the beneficial effect of induced loss mindset intervention was limited to those who were morally committed to non-prejudice. However, this is an important population to consider in encouraging for intervening because previous findings show that those with non-prejudiced self-concept are especially likely to justify their inaction in face of racism through actually derogating the outgroup

#### Appendix A. Participants' demographic questions in Experiment 1 and 2

Lastly, please answer some demographic questions about yourself:

Age \_\_\_\_\_  
Gender \_\_\_\_\_

- o Male (1)
- o Female (2)
- o Other (3) \_\_\_\_\_

Your highest education level:

- o Less than high school (1)
- o High school diploma (2)
- o Bachelor's degree (3)
- o Master's degree (4)
- o PhD (5)
- o Other [not included in analyses]

What is the race or ethnicity which you identify the most with?

- o White or Caucasian (1)
- o Black/African American/Afro-Caribbean (2)
- o Latino/Hispanic (3)
- o Native American (4)
- o Asian (5)
- o Arab (6)
- o Biracial/Mixed: (7) \_\_\_\_\_
- o Other (8) \_\_\_\_\_

(Szekeres, Halperin, Kende, & Saguy, 2019). It remains a direction for future research to conceive messages (most likely outside of the moral mindset domain) that would motivate confronting among those who are less committed to non-prejudice.

#### 4.2. Practical implication

Messages that promote confronting are important because racism is widespread, and while confronting can be effective in decreasing prejudice in perpetrators and bystanders (e.g., Czopp et al., 2006; Czopp & Monteith, 2003), people who witness racism rarely intervene, although they believe they would (e.g., Crosby & Wilson, 2015; Karmali et al., 2017; Kawakami et al., 2009). In the current study, we identified a process and defined messages that can be utilized as a potentially effective intervention tool to increase (some) people's tendency to confront racism, for example in the form of social ad campaigns or workshops. Notably, in Experiment 2, the effect of our intervention seemed to endure across days, and to affect actual confronting behavior. Additionally, due to the feasibility in utilizing the confronting measure, it can be useful for assessment in devising and testing similar interventions in the future.

#### 4.3. Conclusion

Moral courage is a willingness to take a stand in defense of one's own moral principles even when others do not (Miller, 2000; Skitka, 2012). In this research, we tested a way to increase morally courageous behavior and motivate confronting intentions during a situation when people witness racism. We found that exposing people to messages about prospective personal moral failure in regard to not intervening was potentially effective in promoting speaking up against racism. While moral courage is often thought of as a solely altruistic act, we argue that the role of an individual's consideration of their own morality should not be dismissed and can be used to the advantage in encouraging moral behavior in face of racist acts.

What is your religion?

- o No religious affiliation (1)
- o Christian (2)
- o Muslim (3)
- o Jewish (4)
- o Hindu (5)
- o Buddhist (6)
- o Other (7)

What is your socio-political orientation?

Conservative Liberal  
0 1 2 3 4 5 6 7 8 9 10  
*continuous slider*

What is your political affiliation?

- o Democrat (1)
- o Republican (2)
- o Neither (3)
- o Don't want to answer (4)

Compared to other people in your society, what is your economic situation?

- o Wealthy (1)
- o Better than most (2)
- o Good (3)
- o So-So (4)
- o Poor (5)
- o Destitute (6)*coding reversed in study analyses*

#### Appendix B. Moral-prejudice identity self-importance (adapted from Aquino & Reed II, 2002)

Think about a person who is not prejudiced, who is egalitarian and believes that all people are created equal, and who does not discriminate against people based on their gender, sexual orientation, ethnicity or religion. It could be you or it could be someone else. For a moment, visualize in your mind the kind of person who has these characteristics. Imagine how that person would think, feel, and act. When you have a clear image of what this person would be like, answer the following questions.

(9-point scale from 1 = not at all true of me to 9 = completely true of me)  
[*Internalization scale*]

1. It would make me feel good to be a person who has these views and beliefs
2. Being someone who has these views and beliefs is an important part of who I am.
3. I would be ashamed to be a person who has these views and beliefs. (reverse-scored)
4. Having these views and beliefs is not really important to me. (reverse-scored)
5. I strongly desire to have these views and beliefs.

[*Symbolization scale*]

1. The types of things I do in my spare time (e.g., hobbies) clearly identify me as having these views and beliefs.
2. The kinds of books and magazines that I read identify me as having these views and beliefs.
3. The fact that I have these views and beliefs is communicated to others by my membership in certain organizations.
4. I am actively involved in activities that communicate to others that I have these views and beliefs.

From the original scale we excluded the following item: "I often wear clothes that identify me as having these characteristics" (symbolization item).

#### Appendix C. Vignette scenarios and confronting intentions (in Experiment 1)

In this part of the survey, you'll be presented with two ambiguous situations that pose moral dilemmas and you'll be asked questions about it. Please try to place yourself in those described situations as much as possible. Please respond to the questions honestly, according to your own belief, and not according to what you think is expected of you.

[Scenario A]

Imagine you are traveling on the bus, sitting in the back. It's a big bus, only a few people traveling on it. An older teenage boy boards the bus, sits at the back, not far from you, and you hear that he is speaking in Spanish on the phone. Near him, a middle-aged man is sitting, who keeps staring at the boy. The boy hangs up the phone, and the man starts to speak to him. He tells the boy this is the US, and people speak English here. He continues and says that immigrants like him [the boy] should leave this country. Because you believe that this specific boy is treated unfairly, you are debating whether to intervene or not. On the one hand, if you get involved, the man may verbally or even physically attack you. You also don't want to miss your stop which you are approaching soon. If you miss your stop, you'll be late for an important appointment. [Control condition stopped here]

**[Moral loss framing condition continued with the following text]** On the other hand, if you don't get involved, you will probably feel like a bad person. You believe that this action would reveal a bad side of you. That is, you feel that in this situation staying silent means you are behaving immorally. You keep thinking that if you want to avoid moral failure, you should probably intervene. **[Moral gain framing condition continued with the following text]** On the other hand, if you get involved, you will probably feel like a good person. You believe that this action would reveal a good side of you. That is, you feel that behaving morally in this situation means speaking up. You keep thinking that if you want to fulfill your moral ideals, you should intervene.

[Scenario A – Questions]

Based on solely what is described in the text, considering the potential risks involved, how likely it is that in this situation you would perform the following behaviors? (1 = not likely at all to 9 = very much likely)

1. I would stay in my seat and I would not get involved. [reversed]
2. I would quietly leave them and move to the front of the bus. [reversed]
3. I would confront the man and tell him he is racist.
4. I would ask the man to stop assaulting the boy.
5. I would sit next to the boy and start talking to the boy in a friendly manner.
6. I would ask the bus driver to stop the man's behavior.
7. Other suggestion (not mandatory, if you don't write, just mark 1): \_\_\_\_\_

Overall, to what extent you would confront in this situation in order to help the boy?

1 = I would not confront at all to 9 = I would totally confront on a 9-point scale

[Scenario B]

Imagine you are at work, sitting at your desk, working on a difficult assignment. Then you slowly become attentive to a conversation that two of your co-workers are having in the adjacent room. You hear them talking about a third co-worker, who is Muslim. You hear them laughing and making fun of her headscarf, making nasty references about her because of her religion. You don't have many feelings about the mentioned Muslim co-worker, because you hardly know her, but you also don't think it's nice to talk about another individual like that. So, you are debating to confront your co-workers or not. On the one hand, you don't want them to think that you often eavesdrop on their conversations. Additionally, you are working closely with these colleagues, and if they get offended they could even jeopardize your position at work. You also must finish the assignment you are working on as soon as possible. **[Control condition stopped here]** **[Moral loss framing condition continued with the following text]** On the other hand, you are now recalling other unfair situations you've witnessed in the past and how badly you felt about yourself after not confronting. You feel it is your moral obligation to intervene. If you don't intervene, you fail your moral duty, and you may later feel like a worse person morally. You feel you can lose a lot if you don't confront. **[Moral gain framing condition continued with the following text]** On the other hand, you are now recalling other unfair situations you've witnessed in the past and how better you felt about yourself after confronting. You feel it is your moral aspiration to intervene. If you intervene, you succeed to live up to your moral principles, and you may later feel like a better person morally. You feel you can gain a lot if you confront.

[Scenario B – Questions]

Based on solely what is described in the text, considering the potential risks involved, how likely it is that in this situation you would perform the following behaviors? (1 = not likely at all to 9 = very much likely)

1. I would stay in my office and I would not confront them. [reversed]
2. I would sit somewhere else so I don't hear them but I would not confront them. [reversed]
3. I would ask my co-workers to stop insulting her.
4. I would tell my supervisor about my co-workers' conversation.
5. Without being specific, I would just ask them to keep quiet while making sure they know I disprove of their conversation.
6. I would confront my co-workers and tell them they are racists.
7. other suggestion (not mandatory, if you don't write, just mark 1): \_\_\_\_\_

Overall, to what extent you would confront in this situation in order to stand up for her?

1 = I would not confront at all to 9 = I would totally confront on a 9-point scale

#### Appendix D. Moral mindset intervention (in Experiment 2)

Letters are bolded to emphasize the differences between conditions.

[Task 1]

[Loss condition]



[Gain condition]



Please describe what you think this poster means. (please write min. 350 characters)

[Task 2]

Please see a video that was featured in the local news depicting a British woman insulting other passengers she assumed to be Polish or immigrants on a public bus: [Here appeared a video]

[Loss condition]

Later when this incident was reported in the news, a passenger on this bus said that he wished he would have stopped the woman from insulting those other passengers.

Imagine you are this bystander who did not intervene. Please give a short account of your thoughts and feelings.

Start with **"I feel like not intervening revealed a bad side of me..."**

(please write min. 350 characters)

[Gain condition]

Later when this incident was reported in the news, a passenger claimed that he wished he would have stopped the woman from insulting those other passengers.

Imagine you are this bystander who did not intervene. Please give a short account of your thoughts and feelings.

Start with **"I feel like intervening would have revealed a good side of me..."**

(please write min. 350 characters)

[Task 3]

[Loss condition]

Stanislaw Chmielewski (1909–1992) is a Polish Christian man who risked his life to confront injustice and saved a dozen Jews during the Holocaust. He once noted that **not doing what he did would have cost him his moral virtue and he would have felt like a bad person.**

Without knowing more about him, how would you describe Stanislaw's potential thoughts and feelings about his own behavior? (please write min. 350 characters)

[Gain condition]

Stanislaw Chmielewski (1909–1992) is a Polish Christian man who risked his life to confront injustice and saved a dozen Jews during the Holocaust. He once noted that **through this action he gained moral virtue and he feels he became a better person for doing it.**

Without knowing more about him, how would you describe Stanislaw's potential thoughts and feelings about his own behavior? (please write min. 350 characters)

## Appendix E. The Trust Game instructions (in Experiment 2)

*The participants received the following description of the Trust Game:*

You'll be assigned to observe one player and his/her rounds.

Note that this observed player will know you are watching and has an opportunity to send you private messages after each round (we allowed this

option so the observer feels more real to the player). You will have a chance to message him/her back. **Make sure that whatever you say, you do not reveal the purpose of the study to the player you are corresponding with.**

**Pay attention to the game** - 1, the player might message you and 2, it is beneficial for you to get familiar with the game before you are playing it yourself.

### How do you play the Trust Game?

- Two people are in the game. An amount is given to the first player (for example, \$2).
- The first player can decide between three options, to give “ALL” (\$2), “HALF” (\$1), or “NONE” (\$0 = keep all the \$2) of this amount to the other player. Let's say s/he decides to give HALF (\$1).
- This amount is tripled automatically and given to the opponent (\$3).
- The opponent can decide between two options, to “SHARE” the amount with the first player (half it, \$1.50) or give back “NONE” (keep all the \$3). Let's say the opponent decides to SHARE (\$1.50)
- The game ends.
- The first player won \$1 (since s/he kept half of the amount to him/herself in the beginning) plus \$1.50 (which the opponent gave back in the end), ending up in \$2.50.
- The opponent won \$1.50.

Note, players earn the most money if the first player trusts the opponent and gives “ALL”, and the opponent is fair and chooses to “SHARE” (giving half of the earned money back). In this case, following the previous example, they would both end up with \$3.

**Meaning, you can earn the most if you give it ALL and the other player is trustworthy and chooses SHARE.**

## Appendix F. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jesp.2019.103833>.

## References

- Szekeres, H., Halperin, E., Kende, A., & Saguy, T. (2019). *Witnessing, and Not Confronting Racism Amplifies Non-Confronters' Prejudicial Attitudes, (Manuscript under review)*. AND it should be under Swim Hyers reference (due to alphabetical ordering).
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401), 464–477.
- Aoki, J. (2015). *Seeking moral elevation vs. avoiding damnation: An examination of two moral motivational orientations.* (Doctoral dissertation). Retrieved from Lehigh Preserve Database.
- Aquino, K., & Reed, A., II (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83, 1423–1440.
- Batson, C. D. (2011). *Altruism in humans*. USA: Oxford University Press.
- Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry*, 2(2), 107–122.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142.
- Brebels, L., De Cremer, D., Van Dijke, M., & Van Hiel, A. (2011). Fairness as social responsibility: A moral self-regulation account of procedural justice enactment. *British Journal of Management*, 22, S47–S58.
- Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now? *Journal of Experimental Social Psychology*, 56, 110–116.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817–869.
- Crosby, J. R., & Wilson, J. (2015). Let's not, and say we would: Imagined and actual responses to witnessing homophobia. *Journal of Homosexuality*, 62(7), 957–970.
- Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*, 69(2), 117–132.
- 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.
- Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, 90(5), 784.
- Does, S., Derks, B., Ellemers, N., & Scheepers, D. (2012). At the heart of egalitarianism: How morality framing shapes cardiovascular challenge versus threat in Whites. *Social Psychological and Personality Science*, 3(6), 747–753.
- Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology*, 28(1), 94.
- Dutton, D. G., & Lennox, V. L. (1974). Effect of prior “token” compliance on subsequent interracial behavior. *Journal of Personality and Social Psychology*, 29(1), 65.
- Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology*, 45(3), 590–593.
- Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations*, 15(4), 487–502.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Fogelman, E. (n.d.). The rescuer self. Retrieved from: <https://www.yadvashem.org/righteous/resources/the-rescuer-self.html>.
- Hayes, A. F. (2018). PROCESS (version 3.1) [computer software]. Retrieved from <http://www.processmacro.org/download.html>.
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52(12), 1280–1300.
- Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of Personality and Social Psychology*, 66(2), 276.
- Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, 27, 254–263.
- Kaiser, C. R., & Miller, C. T. (2003). Derogating the victim: The interpersonal consequences of blaming events on discrimination. *Group Processes & Intergroup Relations*, 6(3), 227–237.
- Kant, I. (1785). *Groundwork of the metaphysics of morals*.
- Karmali, F., Kawakami, K., & Page-Gould, E. (2017). He said what? Physiological and cognitive responses to imagining and witnessing outgroup racism. *Journal of Experimental Psychology: General*, 146(8), 1073.
- Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science*, 323(5911), 276–278.
- Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology*, 58, 317–344.
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science*, 22(12), 1472–1477.
- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowd-sourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1–10.
- Miller, W. I. (2000). *The mystery of courage*. Cambridge, MS: Harvard University Press.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33.
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, 65(3), 469.
- Nietzsche, F. (1878). *Human, all too human*.
- O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology*, 40(4), 917–946.
- Osswald, S., Greitemeyer, T., Fischer, P., & Frey, D. (2010). Moral prototypes and moral behavior: Specific effects on emotional precursors of moral behavior and on moral behavior by the activation of moral prototypes. *European Journal of Social Psychology*, 40(6), 1078–1094.
- Plant, E. A., & Butz, D. A. (2006). The causes and consequences of an avoidance-focus for interracial interactions. *Personality and Social Psychology Bulletin*, 32(6), 833–846.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811.
- Powell, A. A., Branscombe, N. R., & Schmitt, M. T. (2005). Inequality as ingroup privilege or outgroup disadvantage: The impact of group focus on collective guilt and interracial attitudes. *Personality and Social Psychology Bulletin*, 31(4), 508–521.
- Quinn, K. A., & Olson, J. M. (2011). Regulatory framing and collective action: The interplay of individual self-regulation and group behavior. *Journal of Applied Social Psychology*, 41(10), 2457–2478.
- Sassenberg, K., & Hansen, N. (2007). The impact of regulatory focus on affective

- responses to social discrimination. *European Journal of Social Psychology*, 37(3), 421–444.
- Schaumberg, R. L., & Wiltermuth, S. S. (2014). Desire for a positive moral self-regard exacerbates escalation of commitment to initiatives with prosocial aims. *Organizational Behavior and Human Decision Processes*, 123(2), 110–123.
- Scholer, A. A., Zou, X., Fujita, K., Stroessner, S. J., & Higgins, E. T. (2010). When risk seeking becomes a motivational necessity. *Journal of Personality and Social Psychology*, 99(2), 215.
- Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology*, 65(2), 317.
- Sedikides, C., & Hepper, E. G. (2009). Self-improvement. *Social and Personality Psychology Compass*, 3(6), 899–917.
- Shah, J., & Higgins, E. T. (1997). Expectancy  $\times$  value effects: Regulatory focus as determinant of magnitude and direction. *Journal of Personality and Social Psychology*, 73(3), 447.
- Shelton, J. N., & Stewart, R. E. (2004). Confronting perpetrators of prejudice: The inhibitory effects of social cost. *Psychology of Women Quarterly*, 215–223.
- Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, 4(4), 267–281.
- Skitka, L. J. (2012). Moral convictions and moral courage: Common denominators of good and evil. In M. Mikulincer, & P. R. Shaver (Eds.). *The social psychology of morality: Exploring the causes of good and evil* (pp. 349–365). Washington, D.C: American Psychological Association.
- Skitka, L. J. (2014). The psychological foundations of moral conviction. *Advances in experimental moral psychology* (pp. 148). .
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895.
- Skitka, L. J., & Morgan, G. S. (2014). The social and political implications of moral conviction. *Political Psychology*, 35, 95–110.
- 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.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90(2), 245.
- Winterich, K. P., Aquino, K., Mittal, V., & Swartz, R. (2013). When moral identity symbolization motivates prosocial behavior: The role of recognition and moral identity internalization. *Journal of Applied Psychology*, 98(5), 759.
- Winterich, K. P., Mittal, V., & Aquino, K. (2013). When does recognition increase charitable behavior? Toward a moral identity-based model. *Journal of Marketing*, 77(3), 121–134.
- Zaal, M. P., Laar, C. V., Ståhl, T., Ellemers, N., & Derks, B. (2011). By any means necessary: The effects of regulatory focus and moral conviction on hostile and benevolent forms of collective action. *British Journal of Social Psychology*, 50(4), 670–689.
- Zaal, M. P., Van Laar, C., Ståhl, T., Ellemers, N., & Derks, B. (2012). Social change as an important goal or likely outcome: How regulatory focus affects commitment to collective action. *British Journal of Social Psychology*, 51(1), 93–110.