EÖTVÖS LORÁND UNIVERSITY FACULTY OF EDUCATION AND PSYCHOLOGY

Doctoral Dissertation Summary

Szekeres Hanna

THE FAILURE TO CONFRONT PREJUDICE

Doctoral School of Psychology

Head of School: Prof. Dr. Anikó Zsolnai, Eötvös Loránd University

Programme name: Socialization and Psychology of Social Processes Program

Head of Programme: Dr. Nguyen Luu Lan Anh, Eötvös Loránd University

Supervisors:

Dr. Anna Kende, Eötvös Loránd University

Dr. Tamar Saguy, Interdisciplinary Center (IDC) Herzliya

Prof. Dr. Eran Halperin, Hebrew University Jerusalem

Budapest, 2021

Doctoral candidate's list of publications

Publications cited in the dissertation:

Szekeres, H., Halperin, E., Kende, A., & Saguy, T. (2019). The effect of moral loss and gain mindset on confronting racism. *Journal of Experimental Social Psychology, 84*, 103833.

Szekeres, H., Shuman, E., & Saguy, T. (2020). Views of sexual assault following# MeToo: The role of gender and individual differences. *Personality and Individual Differences, 166*, 110203.

Szekeres, H. (2020). Kedvelni vagy tisztelni? Az előítéletek csökkentése a melegszívűség és kompetencia dimenzióiban. *Alkalmazott Pszichológia, 20*(4), 123-157.

Saguy, T., & **Szekeres**, H. (2018). Changing Minds via Collective Action: Exposure to the 2017 Women's March Predicts Over-time Decrease in (Some) Men's Gender System Justification. *Group Processes & Intergroup Relations, 21(5*), 678-689.

Additional publications:

Saguy, T., **Szekeres**, H., Nouri, R., Goldenberg, A., Doron, G., Dovidio, J. F., ... & Halperin, E. (2015). Awareness of intergroup help can rehumanize the out-group. *Social Psychological and Personality Science, 6*(5), 551-558.

Bruneau, E., **Szekeres**, H., Kteily, N., Tropp, L. R., & Kende, A. (2020). Beyond dislike: Blatant dehumanization predicts teacher discrimination even (and especially) among teachers low in prejudice. *Group Processes and Intergroup Relations*, *23*(4), 560-577.

I. Introduction

Historical events of the 20th century, such as the Holocaust in Europe, genocide in Rwanda, and massacre in Srebrenica, prompted longstanding societal and empirical interest in the phenomenon of people's failure to stand up for others and intervene in times of racial or ethnic atrocities. While currently in today's democratic societies such blatant mass tragedies do not transpire, structural and everyday forms of racism still frequently occur and can quickly intensify. For example, consider recent events in this decade when prejudicial sentiment spiraled and lead to the White supremacist "Unite the Right ally" events in 2017 in Charlottesville (in the U.S.), the anti-immigrant wave of attacks in the aftermath of Brexit in 2016 (in the UK), or the serial murders perpetrated by Neo-Nazis against people of Roma ethnicity in 2008-2009 (in Hungary).

In our everyday lives, from time to time, we witness someone expressing prejudice on the bus, at the grocery store, at school, or at the workplace. From time to time, we witness people being discriminated against because of their religion, the color of their skin, or who they choose as their romantic partner. How do we react in these instances? Do we say something? Imagine your co-worker voicing something prejudicial and hurtful about another co-worker who is a minority individual, and then also discriminates them by not inviting them to join an important project at work. In this instance, you may feel the words coming up from your stomach, in your lung, in your throat – but in the end you decide not to say anything.

While people generally believe they would stand up for others and act against injustice, prejudice and discrimination, in reality they often fail to do so (e.g., Crosby & Wilson, 2015; Kawakami et al., 2009). In my dissertation research, I investigated the psychological motivations and consequences of witnessing and (failure of) confronting prejudice and discrimination. In the focus of my research is the bystander, who is *not* the member of the stigmatized group, but has an opportunity to confront the source of prejudice. We conducted our experiments in two countries, in the United States and in Hungary, across various intergroup contexts, where the outgroup was a racial, ethnic or religious minority.

A. Prevalence of prejudice and its negative impact

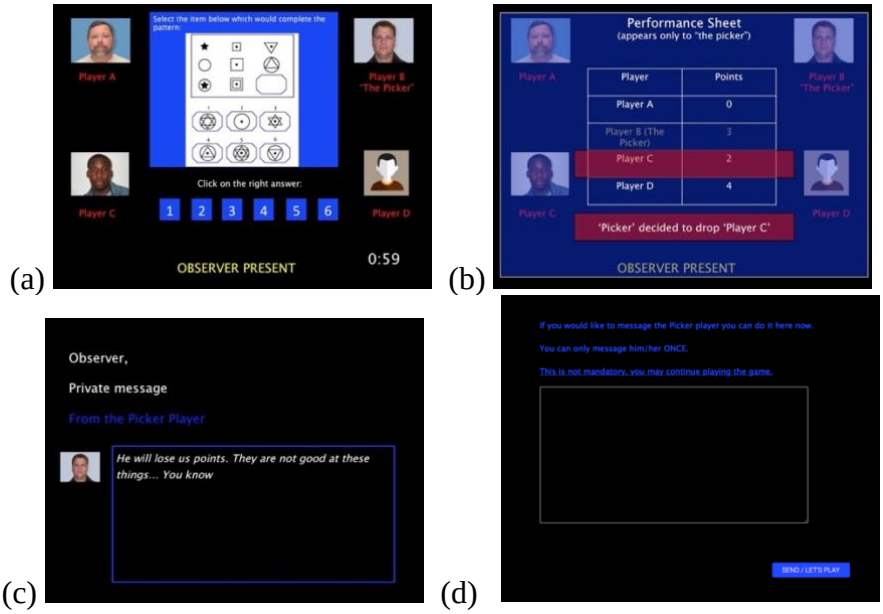
Both in the US and in Hungary, discrimination based on race, ethnicity or religion is against the law, yet discrimination of minorities occurs in various areas of their lives, such as in housing, access to health care, employment, education, law-enforcement, or jurisdiction (e.g., Lee et al., 2019 for US; Sik & Simonovits, 2012 for Hungary). Moreover, minorities in

both countries frequently experience "everyday prejudice", such as staring, prejudicial slurs, insensitive jokes or microagressions in the form of political discourse (e.g., through media) to interpersonal interactions (e.g., in public areas, workplace). Experiencing such prejudice take a toll on the stigmatized individuals in various ways (e.g., Swim et al., 2003; Sue et al., 2007). For one, prejudice and discrimination in employment and in workplace setting affects hiring, and one's professional ambition, advancement, and job satisfaction (e.g., Triana et al., 2015) – thus it has an economic toll on the stigmatized individuals. Additionally, exposure to prejudice has a psychological toll on the person (for review see Barreto & Ellemers, 2015), causing lower self-esteem and self-worth (e.g., Twenge & Crocker, 2002), which not only negatively impacts educational and professional performance and achievement (e.g., Nguyen & Ryan, 2008; Walton & Cohen, 2007), but even threats mental and physical well-being and health (e.g., Paradies et al., 2015; Schmitt et al., 2014). These negative impacts highlight the importance of focusing on methods to negate prejudice expression even in its "everyday" form. One such strategy is confrontation of people who openly espouse prejudice.

B. Overview of the present research and its findings

In the current research, across five-seven experiments, I investigated (1) the self-justifying harmful consequences of bystanders' inaction on their own intergroup attitudes (research project #1), and related (2) psychological–moral messages that could promote speaking up in face of prejudice (research project #2). We conducted online experiments in the US and in Hungary (N = 1629), in various intergroup contexts where the outgroup minority was either African American, Muslim American or Latino (US), or Jewish (Hungary). For the current research, to test and allow for actual confronting, I developed and programmed myself and used across studies (Pilot, Studies 1-3 and Study 5) an **online behavioral paradigm**, where participants believed they are witnessing prejudice and have an opportunity to confront. In this paradigm, participants observed and played an online game, during which an observed player behaved discriminatively towards another player who was an outgroup minority and then privately messaged the participant with a prejudiced statement, and the participant had an opportunity to reply. For scenes from the game paradigms, see Figure 1 and Figure 2.

Figure 1. Scenes from the Logic-IQ game paradigm used in the Pilot studies.



(a) (b) (c) (d)

Note. (a) During a question posed to players; (b) Performance sheet with players' earned points and showing that Black player is eliminated by the prejudiced (Picker) player; (c) Picker player's prejudiced message; (d) Message box providing an opportunity to respond to the prejudiced (Picker) player.

Figure 2. Scenes from the Trust/Share game paradigm used in Studies 1-3 and Study 5.

(a) (b)

Note. (a) Prejudiced player (Mark) is playing with the Muslim player (Hakim) and denies him money; (b) Mark messages the participant with a prejudiced remark about Muslims.

In research project #1 (Pilot and studies 1-3, N = 922), we tested the impact of failure to confront prejudice on the intergroup attitudes of the bystander. Drawing on cognitive dissonance and self-justification theories, we predicted and found that those who did not confront the prejudiced perpetrator, albeit having an opportunity to, would subsequently endorsed more negative attitudes about the outgroup compared to their initial attitudes, and to control groups – likely in order to justify and reconcile with prior inaction. As an effect of inaction, they were also more likely to trivialize the prejudiced incident and deny responsibility for intervening. In this work, we demonstrated a route via which prejudice (not confronted) can intensify in a given social environment.

In research project #2, conducted in the U.S., we explored the impact of moral mindsets on bystanders (Studies 4-5, N = 707). Specifically, we tested whether the prospect of moral

loss (failure) or moral gain (success) in relation to intervening can motivate people to confront prejudice. Drawing on regulatory focus and prospect (loss aversion) theories, we predicted and found that a moral loss framing/mindset increases confronting intentions among those who are morally committed to non-prejudice (possibly due to a desire to safeguard their moral self-concept). That is, if a person cares about being non-prejudiced, the prospect of loss to one's moral self-regard if action is not taken, can actually motivate acting the desired behavior. Meanwhile a moral gain mindset had no effect on confronting. In this research, we devised a moral mindset intervention and messages that (even a few days later) affected actual real-life confronting, which thus can be effectively used in promoting people's tendency to speak up against prejudice, or can be tailored to other morally desired behavior.

II. Research Project #1: The Motivated Prejudice Effect – Endorsing Negative Intergroup Attitudes to Justify Not Confronting Prejudice (Studies 1-3)

A. Theoretical background, rationale and hypotheses

On many occasions, people may feel upset when witnessing prejudice (e.g., Schmader et al., 2012; Torres et al., 2019), but nevertheless may not act against it. For example, heterosexual participants who imagined witnessing a homophobic slur reported higher intentions of confronting than people who actually witnessed the slur (Crosby & Wilson, 2015). Similarly, even though White Americans anticipated acting against someone who expressed racism, those put in that actual situation forewent punishing the perpetrator (Karmali et al., 2017; Kawakami et al., 2009). However, people do not particularly feel reconciled with failing to condemn blatant injustice and racism, and they may experience some psychological discomfort. For example, when White Americans felt they should not behave in a prejudiced manner towards minorities but were made to consider how they might do so, experienced discomfort (Voils et al., 2002; Zuwerink et al., 1996). Also, heterosexuals who recalled witnessing and not confronting anti-gay bias reported to feel more conflicted with their inaction than those who confronted (Dickter, 2012). To resolve such discomfort, people may engage in self-justifying rationalization. Although investigating the *target* group, but Rasinski and colleagues (2013) notably tested reactions following inaction to bias. They found that female participants who initially valued confronting and were given the opportunity to confront, but did not, made more favorable evaluations of the sexist perpetrator, compared to those who had no chance to confront, and they also devalued confronting socially inappropriate behavior in

general – researchers suggested this was possibly so to reduce dissonance for inaction (Rasinski et al., 2013; see also Mallett et al., 2019, Study 2).

Prior research primarily focused on how the negative societal consequence of not confronting prejudice is the failure to challenge the perpetrator. Indeed, confronting prejudice can be effective in changing perpetrators' beliefs and reduce prejudice (e.g., Burns & Monteith, 2019; Chaney & Sanchez, 2018; Czopp, Monteith, & Mark, 2006). In the current research, by applying the logic of self-justification to those bystanders who do not belong to the stigmatized outgroup, and by moving the focus beyond the perpetrator, we tested a process where as an effect of inaction bystanders come to endorse more prejudice, thereby rendering the problem of inaction even more socially problematic.

Based on the theory of *cognitive dissonance reduction* (Abelson et al., 1968; Festinger, 1957), I expected that people who witness prejudice and do not confront (while given an opportunity to) would be motivated to resolve the inconsistency between their attitudes (about condemning racial injustice) and their inaction (in face of injustice) by changing their attitudes and appraisal about how unjustified was the witnessed incident. We propose that those who fail to confront would rationalize that the outgroup deserved the treatment, and the observed prejudice was based on a reasonable judgment and has a kernel of truth in it (i.e., along the line of "after all, they are kind of like that"). Thus, not confronting prejudice would reinforce and increase the witnessed negative attitudes about the outgroup. Beyond this attitude change, based on strategies that people often employ to reduce cognitive dissonance (see McGrath, 2017), we also tested two additional dissonance reduction stratetgies and outcomes of inaction: trivialization (Festinger, 1957; Simon et al., 1995) of the prejudiced incident (i.e., "what happened wasn't even that harmful") and denial of responsibility (Gosling et al., 2006) for intervening (i.e., "It was not my responsibility to do something").

B. Methods

To test our predictions about the proposed motivated prejudice effect, I used the self-developped behavioral paradigm presented above in order to place participants in a real situation where they can choose to confront or not (in the main experimental condition). Initially, we conducted two pilot experiments,[1] where we tested and found that White American participants who witnessed but did not confront prejudice against a Black player, had subsequently more negative outgroup attitudes and trivialized the prejudiced incident more

[1] Actually our first study was an in-lab situational and interview study in Israel (n=11).

than those who did not witness prejudice, or witnessed the same prejudice but had no opportunity to confront (control groups, and in analyses we controlled for participants' baseline socio-political attitudes also). While the pilot studies provided initial support for our predictions, they had a limitation of lacking pre-posttesting.[2]

Following the pilot studies and pre-tests of the game, across three experiments (studies 1-3), I used a mixed within- and between-subjects design, where I assessed participants both prior and following the incident (pre-posttest). This design enabled us to test overtime changes in outgroup attitudes among those who did not confront, and to compare those changes to control groups, in order to show that people *become* more prejudiced as a function of not confronting. To increase external validity of the research, studies were conducted in different countries and intergroup contexts, in Hungary with Jewish outgroup (Study 1, N = 138) and in the U.S. with Muslim (Study 2, N = 120), and Latino outgroup (Study 3, N = 410). The witnessed incident and measures were framed around intergroup trust and liking, however the actual slurs varied to fit the predominant prejudice about the target outgroup, and outgroup attitude measures varied accordingly (mainly in study 1) and also by keeping in mind what is used in the literature (mainly in study 2-3).

In **studies 1–2**, in the control condition, participants observed another type of prejudice not rooted in intergroup membership (but "interpersonal") and had an opportunity to react. We predicted no intergroup attitude change for those who did not confront interpersonal bias, compared to intergroup bias. This would show that the proposed effect is not specific to a personality type who does not confront socially inappropriate behavior in general (who is not assertive), nor is it a derogatory response resulting from a deflated self-esteem (Fein & Spencer, 1997) that would be brought upon by *any* personal failure of not confronting (hereafter "personal" account).

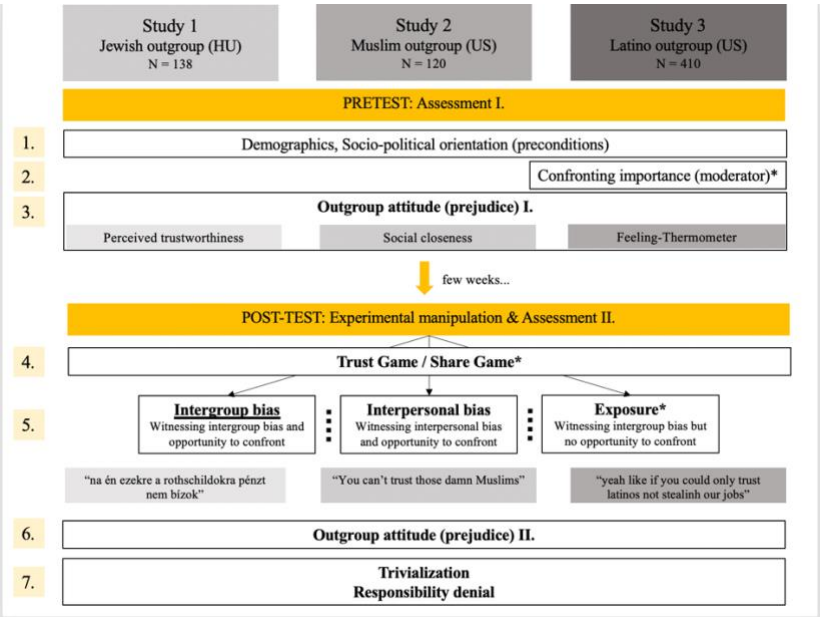
In **Study 3**, we added another control condition beyond interpersonal, where participants were only exposed to intergroup prejudice (exposure condition), that is, observed the exact same intergroup prejudice but did not have an opportunity to reply to the message, aka. confront. Since these participants have external justification (and no personal responsibility) for staying silent, they will not engage in attitude change for justifying their inaction. If there is no attitude change following only exposure (compared to intergroup non-

[2] There is an inherent methodology problem (selection issue) in our research question about those who do not confront, given we only analyse these participants from the experimental condition (for a similar approach see Rasinski et al., 2013). Besides our methodological attempts to flatten confronting rate, the pre-posttest design is the most optimal approach to address this issue.

confronters) then we can rule out desensitization, persuasion or change in normative context (e.g., Blanchard et al., 1991; 1994) or victim blaming triggered by just-world beliefs (e.g., Lerner & Simmons, 1966) as alternative explanations for our proposed effect (hereafter "exposure" account).

In all studies, study procedure was alike, (and all online) see Figure 3. Participants first completed a pre-test assessment of outgroup attitudes (prejudice) and socio-political orientations (e.g., political ideology, social dominance orientation).[3] Few weeks later, not knowing the studies are connected, participants returned to a post-test where they participated in the game, and were randomly assigned to intergroup bias or interpersonal bias, or exposure condition (only in Study 3). Following the game, in a seemingly unrelated survey among filler scales participants responded again to outgroup attitude measure (post-test assessment). At the end of the experimental session, across all conditions, we assessed (cunningly) trivialization of the intergroup prejudiced event and responsibility denial for intervening. Data and analyses of studies 1-3, and the pre-registration of Study 3 is available here: https://osf.io/36ay8/?view_only=2f41a047b78b46e99055e5255a558336.

Figure 3. Study procedures across studies 1-3.



Note. * Only in Study 3.

[3] Across all studies, we tested and found no significant differences between groups in *baseline* (measured at pre-tes) outgroup attitudes or socio-political orientations (or on perceived importance of confronting prejudice, Study 3), suggesting that intergroup non-confronters are not characteristically more conservative, non-egalitarian or prejudicial than the control groups.

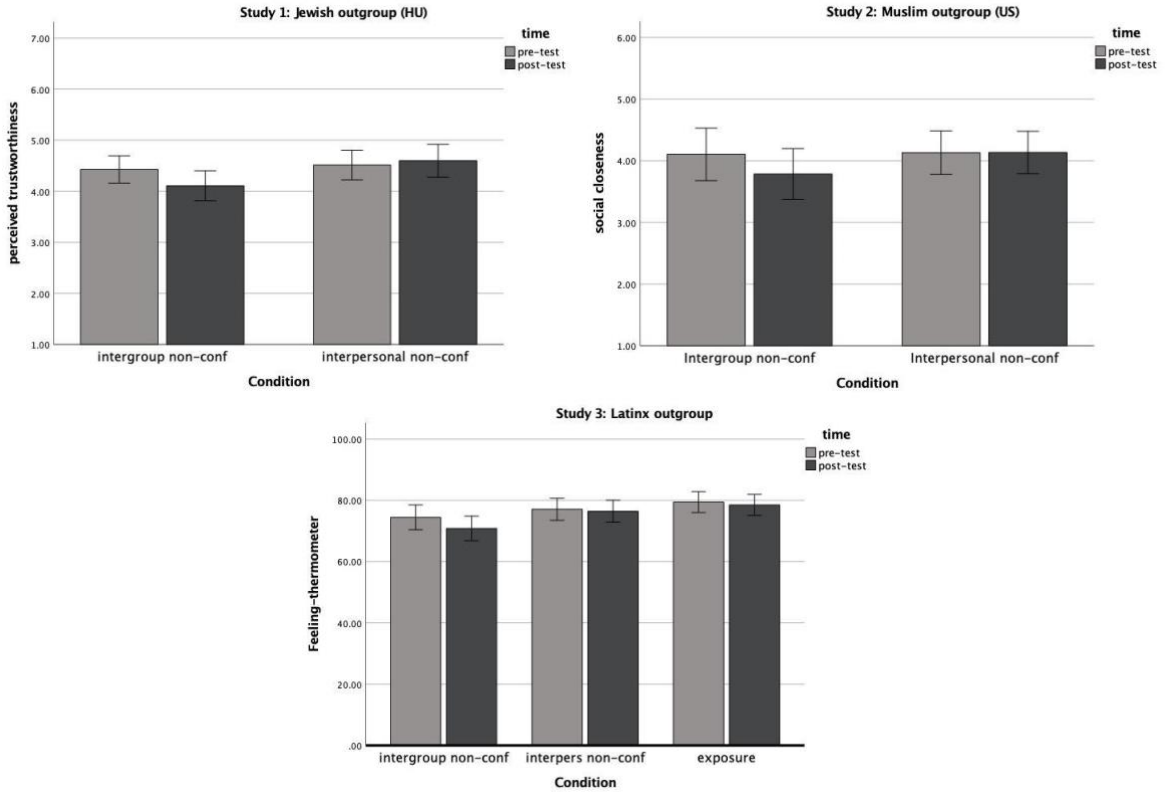
C. Results

According to predictions, we found that those who witnessed intergroup prejudice and had an opportunity to confront the perpetrator, but did not, endorsed significantly more negative outgroup attitudes relative to their own attitudes prior to the incident (see Figure 4). There were no overtime attitude changes within control groups.[4] Moreover, following the incident (and not before), intergroup non-confronters showed more negative outgroup attitudes, trivialization and responsibility denial than (1) those who witnessed and did not confront other non-intergroup type bias (i.e., interpersonal), thus ruling out the "personal" accounts; And (2) those who witnessed the same intergroup prejudice scenario but did not have an opportunity to confront, thus ruling out the "exposure" account (see Figure 5). (Internal meta-analyses supported the robustness or consistency of these findings.) That is, witnessing prejudice alone, without opportunity to confront (and thus without need for justifying prior inaction), did not normalize prejudice endorsement. We suggest that our findings are explained through a dissonance-induced self-justification mechanism (Abelson et al., 1968; Festinger, 1957; Rasinski et al., 2013), whereby people felt an inconsistency between their cognition (prejudice is wrong and should be contested) and their inaction in face of prejudice, and they were motivated to reduce this dissonance.[5]

[4] Nor among intergroup confronters, ruling out attitude polarization (e.g., Saguy & Szekeres, 2018) as alternative explanation.

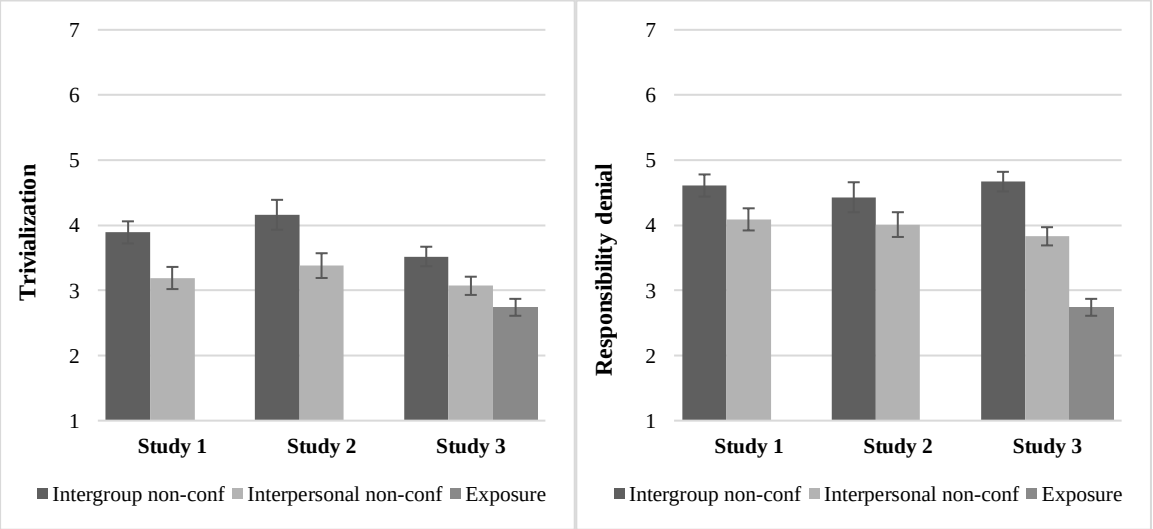
[5] In Study 3, we tested and found that the change in attitudes among intergroup non-confronters was (somewhat) dependent on participants' baseline perceived importance of confronting prejudice. While the (moderation) interaction was not significant, simple effect analyses indicated that those lower on valuing confronting did not show the proposed effect – presumably because inaction did not contradict their personal values, thus they did not seek justification for not confronting. This partial evidence is important because it rules out self-perception theory (Bem, 1967) as plausible explanation for our proposed effect (because it would suggest that the effect would be strongest exactly among those holding weak attitude).

Figure 4. Outgroup attitudes as a function of experimental groups and pre- and post-test sessions (time) across studies 1-3. (Error bars *95% CI* indicated)



Note. The results of the internal meta-analyses for (1) the mixed ANOVA (time X group) interaction effect: M*r* = 0.12, *Z* = 3.15, $p < .005$, two-tailed; (2) the repeated measures simple effect within intergroup non-confronters: M*r* = 0.18, *Z* = 2.67, $p < .01$, two-tailed; And (3) ANCOVA between-groups comparison (where covariate is pre-test prejudice scores): M*r* = 0.13, *Z* = 3.22, $p < .005$, two-tailed.

Figure 5. Trivialization of the intergroup prejudiced event and responsibility denial for intervening as a function of experimental groups across studies. (Stander Error bars are indicated).



Note. The results of the internal meta-analyses for ANCOVA between-groups comparison (where covariate is pre-test prejudice scores): (1) Trivialization: M*r* = 0.19, *Z* = 5.03, *p* < .0001, two-tailed; And (2) Ressponsibility denial: M*r* = 0.17, *Z* = 4.37, *p* < .0001, two-tailed.

III. Research Project #2: When it's your loss – The effect of moral loss and gain mindset on confronting prejudice (Studies 4-5)

A. Theoretical background, rationale and hypotheses

Based on the findings of research project #1, considering how people likely feel some discomfort following their failure to confront prejudice, I considered potential ways to use *anticipated* intrapersonal moral costs to actually motivate confronting. Maintaining a non-prejudiced self-image is important to many individuals (e.g., Plant & Butz, 2006). 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), inhibiting prejudiced responses to jokes (Monteith, 1993), or being more generous to an outgroup member (Dutton & Lake, 1973). In a situation, when people witness prejudice and have an opportunity to confront, they may equally experience their moral identity threatened or questioned, and decide to confront to avoid a potential loss to their moral self-concept.

Based on Tversky and Kahneman's (1991, 1992) prospect theory, losses inflict psychological harm to a greater degree than gains gratify, which means that people are more willing to run risks to avoid or recoup losses than to make gains. This suggests that the psychological costs of falling short of one's moral self-concept should be a stronger motivating

force than equivalent psychological gains of fulfilling one's moral ideals. Accordingly, I propose that the prospect of *feeling immoral* as a result of omission to confront prejudice would be stronger in motivating confronting, than the prospect of *feeling moral* after confronting. I further assert that this loss aversion effect is contingent upon the individual's moral commitment to non-prejudice. That is, a moral loss mindset would only activate loss aversion among those who to begin with possess a strong non-prejudiced moral self-concept, which then could be at stake during witnessing prejudice.[6]

At the same time, 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. Instead, for them, a focus on gains to one's moral self-concept could drive more confronting because it is seen as an opportunity to improve moral self-regard (see self-improvement; Leary, 2007; Sedikides & Hepper, 2009). Also they may confront to gain moral credits prospectively in the domain of prejudice (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 et al., 2009).

Taken together, we predicted that participants' moral commitment to non-prejudice would moderate the effects of moral mindset on confronting prejudice. Specifically, a moral loss (vs. control) mindset would significantly increase confronting tendencies among those strongly morally committed to non-prejudice (possibly to safeguard their moral self-concept), 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 (possibly to enhance their moral self-concept), and would not affect those strongly committed (H2).

B. Methods

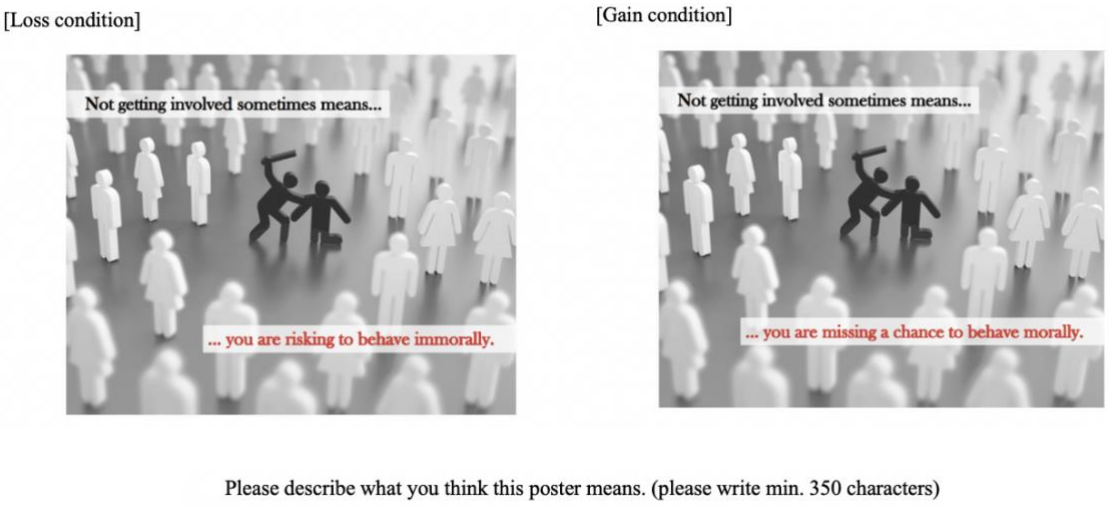
We tested our hypotheses in two online experiments on U.S. participants. In **Study 4**, we manipulated moral mindset and measured the self-reported tendency to confront prejudice. As hypothesized moderating variable, participants responded to scales measuring the strength

[6] Research on self-regulatory focus and collective action supports this possibility. It was found that individuals under prevention focus (corresponding to the loss mindset) are more likely to engage in action aimed at amending injustice directed towards their own group (although this was not related to witnessing prejudice), 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).

of their moral commitment to non-prejudice (Moral Conviction, Skitka & Morgan, 2014; Moral Identity, Aquino & Reed, 2002; Both adapted to prejudice; e.g., "standing up against prejudice is a reflection of my core moral beliefs"). All participants were provided with two vignettes, each depicting an everyday scenario of prejudice (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 moral self-concept loss, moral self-concept gain or neither. In this latter control group participants were exposed to the same scenarios, also involving some non-morality related costs to confronting – to rule out a morality priming effect account (Osswald et al., 2010). Then we asked participants about their willingness to confront in the situations described.

In **Study 5**, we used a different manipulation where participants might internalize the moral messages more, and used the previously developped Trust Game paradigm to measure actual confronting behavior. Participants first filled out the scales of moral commitment to non-prejudice. Then they were randomly assigned either to a moral loss or moral gain mindset intervention, or to an empty control condition. The online intervention, which I designed for the current research consisted of three engaging tasks, where participants had to respond (in writing) to visual and textual materials (for an example task see Figure 6). After a couple of days, we approached the same participants with the Trust Game, where they witnessed prejudice against Muslims and had an opportunity to confront the prejudiced player.

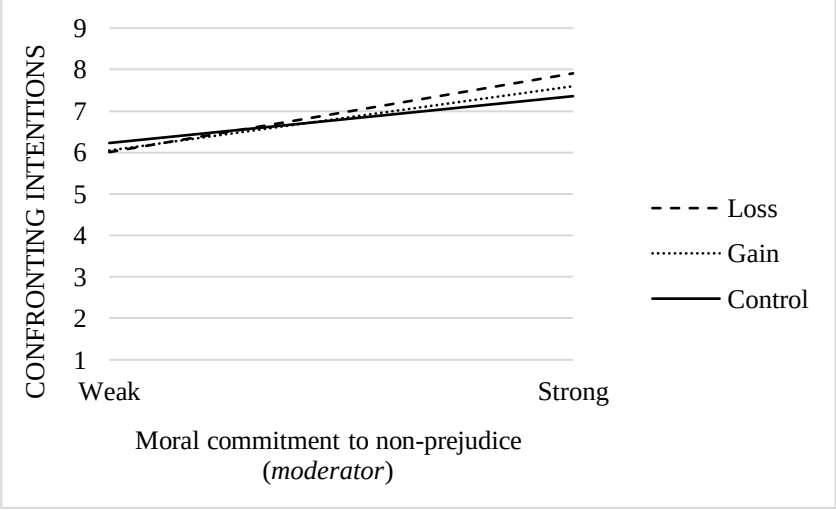
Figure 6. A task in the moral mindset intervention.



C. Results

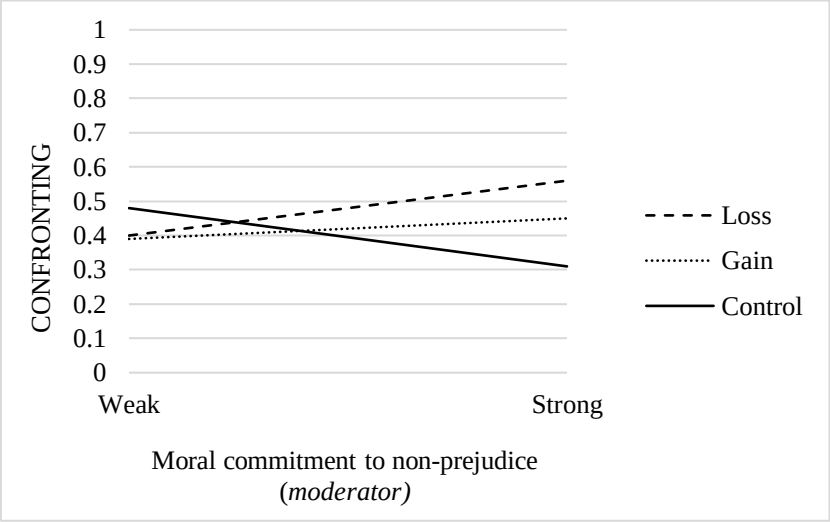
The studies provide partial support to our predictions. First, we failed to find support for the hypothesis regarding the moral gain mindset (H2). This mindset did not seem to be effective in increasing confronting rate in comparison to the control group at any level of moral commitment to non-prejudice. Regarding our other hypothesis (H1), we predicted and found evidence that moral framing can affect the tendency to confront racism, and this is dependent on participants' non-prejudiced moral commitment. Among those with high moral commitment to non-prejudice, a loss mindset led to more confronting, compared to the control condition (H1). See Figure 7 and 8. Likely, the loss framing activated motivation to safeguard one's moral non-prejudiced self-concept (Dutton & Lennox, 1974; Monteith, 1993).[7]

Figure 7. Confrontating intentions (on a 9-point scale) as a factor of mindset framing condition (loss vs. gain vs. control) and participants' moral commitment to non-prejudice (moral conviction) in Study 4.



[7] However, as a limitation, in the two studies different morality scales moderated the found effect.

Figure 8. Confronting (yes or no; visualizing probabilities) as a factor of mindset framing condition (loss vs. gain vs. control) and participants' moral commitment to non-prejudice (MID-symbolization) in Study 5.



IV. Discussion

In the current research (RP#1), I found that when bystanders choose not to confront prejudice, albeit having an opportunity to, they come to endorse more negative intergroup attitudes (possibly in order to justify their inaction). Such attitude change as a factor of failure to confront are harmful as they build tolerance for prejudicial atrocities in the long-run which in turn likely to go uncontested. Thereby creating a destructive cycle, where prejudice not confronted exponentially amplifies in a given social environment. Being aware of how intrapersonal costs of not confronting can lead to intergroup costs, I considered potential ways to use *anticipated* intrapersonal costs to actually motivate confronting. Accordingly, I found (RP#2) that if a person cares about being non-prejudiced, messages about the potential loss of one's sense of morality if action is not taken can actually trigger intervening. In this research, I developped messages and an intervention tool that can be utilized to increase (some) people's tendency to confront prejudice (and potentially even other forms of immoral behavior). This moral mindset message could be implemented in the field in the form of social ad campaigns or workshops or used by civil organizations for developing tools for promoting tolerance, and applied in companies or schools, where the community is diverse, and instances of bias can readily occur and are likely to go uncontested.

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. On that note, a major strength of my research is the employment of the online behavioral paradigm. This way I was

able to place participants in an allegedly real situation, and measure actual confronting behavior. While there are many benefits of conducting the research online (recruitting larger samples across various contexts), the generalizability of our findings (external validity) is limited in regard to confrontation that occurs in face-to-face interactions. Yet, overall, we assume that both investigated phenomena are driven by psychological processes that are not specific to online contexts and overall has relevance to other naturalistic forms of social interactions.

Given the growth of diverse societies and simultaneous and occasional rise in prejudice (Craig & Richeson, 2014), potential bystanders to prejudice are becoming increasingly common, rendering the focus of the present research timely and relevant. Confronting prejudice is an important socio-political behavior because it provides an opportunity to communicate disagreement or disapproval with prejudicial treatment within an interpersonal interaction, and to promote an inclusive climate (Mallett & Monteith, 2019).

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (Eds.). (1968). Theories of cognitive consistency: a sourcebook. Rand-McNally: Chicago.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423-1440.

Barreto, M., & Ellemers, N. (2015). Detecting and experiencing prejudice: New answers to old questions. *Advances in experimental social psychology*, *52*, 139-219.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183-200.

Blanchard, F. A., Lilly, T., & Vaughn, L. A. (1991). Reducing the expression of racial prejudice. *Psychological Science*, *2*(2), 101-105.

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993.

Burns, M. D., & Monteith, M. J. (2019). Confronting stereotypic biases: Does internal versus external motivational framing matter?. *Group Processes & Intergroup Relations*, *22*(7), 930-946.

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.

Chaney, K. E., & Sanchez, D. T. (2018). The endurance of interpersonal confrontations as a prejudice reduction strategy. *Personality and Social Psychology Bulletin*, *44*(3), 418-429.

Craig, M. A., & Richeson, J. A. (2014). More diverse yet less tolerant? How the increasingly diverse racial landscape affects white Americans' racial attitudes. *Personality and Social Psychology Bulletin*, *40*(6), 750-761.

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.

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.

Dickter, C. L. (2012). Confronting hate: Heterosexuals' responses to anti-gay comments. *Journal of Homosexuality*, *59*(8), 1113-1130.

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.

Fein, S., & Spencer, S.J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, *73*(1), 31.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Gosling, P., Denizeau, M., & Oberlé, D. (2006). Denial of responsibility: a new mode of dissonance reduction. *Journal of Personality and Social Psychology*, *90*(5), 722.

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.

Lee RT, Perez AD, Boykin CM, Mendoza-Denton R (2019) On the prevalence of racial discrimination in the United States. *PLoS ONE, 14*(1).

Lerner, M. J., & Simmons, C.H. (1966). Observer's reaction to the" innocent victim": Compassion or rejection?. *Journal of Personality and Social Psychology*, *4*(2), 203.

Mallet, R. K., & Monteith, M. J. (Eds.) (2019). *Confronting prejudice and discrimination: The science of changing minds and behaviors*. London, UK: Academic Press.

Mallett, R. K., Ford, T. E., & Woodzicka, J. A. (2019). Ignoring sexism increases women's tolerance of sexual harassment. *Self and Identity*, 1-17.

McGrath, A. (2017). Dealing with dissonance: A review of cognitive dissonance reduction. *Social and Personality Psychology Compass*, *11*(12).

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.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314.

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.

Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., ... & Gee, G. (2015). Racism as a determinant of health: a systematic review and meta-analysis. *PloS one*, *10*(9), e0138511.

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.

Rasinski, H. M., Geers, A. L., & Czopp, A. M. (2013). "I guess what he said wasn't that bad" dissonance in nonconfronting targets of prejudice. *Personality and Social Psychology Bulletin*, *39*(7), 856-869.

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.

Schmader, T., Croft, A., Scarnier, M., Lickel, B., & Mendes, W. B. (2012). Implicit and explicit emotional reactions to witnessing prejudice. *Group Processes & Intergroup Relations*, *15*(3), 379-392.

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: a meta-analytic review. *Psychological Bulletin*, *140*(4), 921.

Sedikides, C., & Hepper, E. G. (2009). Self-improvement. *Social and Personality Psychology Compass*, *3*(6), 899-917.

Sik, E. & Simonovits, B. (2012). *A diszkrimináció mérése. e-tankönyv*. Budapest, ELTE. Retrieved from: https://www.tarki.hu/hu/about/staff/sb/Diszkriminacio_merese.pdf

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: the forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, *68*(2), 247.

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., Cohen, L. L., Fitzgerald, D. C., & Bylsma, W. H. (2003). African American college students' experiences with everyday racism: Characteristics of and responses to these incidents. *Journal of Black psychology*, *29*(1), 38-67.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: implications for clinical practice. *American psychologist*, *62*(4), 271.

Triana, M. D. C., Jayasinghe, M., & Pieper, J. R. (2015). Perceived workplace racial discrimination and its correlates: A meta-analysis. *Journal of Organizational Behavior*, *36*(4), 491-513.

Torres, L., Reveles, A. K., Mata-Greve, F., Schwartz, S., & Domenech Rodriguez, M. M. (2020). Reactions to Witnessing Ethnic Microaggressions: An Experimental Study. *Journal of Social and Clinical Psychology*, *39*(2), 141-164.

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.

Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing whites, blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin*, *128*, 371–408.

Voils, C. I., Ashburn-Nardo, L., & Monteith, M.J. (2002). Evidence of prejudice-related conflict and associated affect beyond the college setting. *Group Processes & Intergroup Relations*, *5*(1), 19-33.

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.

Zuwerink, J.Z., Devine, P.G., Monteith, M. J., & Cook, D.A. (1996). Prejudice toward Blacks: With and without compunction?. *Basic and Applied Social Psychology*, *18*(2), 131-150.
Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of personality and social psychology*, *92*(1), 82.