The Impact of Affiliative Motivation and Trust on Confrontations of Racism

Jamie Patrianakos

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Social Psychology Commons

Recommended Citation
https://ecommons.luc.edu/luc_diss/3897

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2021 Jamie Patrianakos
LOYOLA UNIVERSITY CHICAGO

THE IMPACT OF AFFILIATIVE MOTIVATION AND TRUST ON CONFRONTATIONS OF RACISM

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

PROGRAM IN SOCIAL PSYCHOLOGY

BY

JAMIE L. PATRIANAKOS

CHICAGO, IL

AUGUST 2021
Copyright by Jamie Patrianakos, 2021
All right reserved.
ACKNOWLEDGEMENTS

My academic achievements would not have been possible without the unwavering support of those around me. To Dr. Robyn Mallett, my mentor and dissertation chair, thank you for providing me with all the tools I needed to finish this marathon and make it to the light at the end of this tunnel. You are as persistent in using metaphors about graduate school as you are in your compassion for your students. I am truly grateful to have you in my corner. I would like to thank my other committee members, Dr. Tracy DeHart, Dr. Yael Granot, and Dr. Matthew Miller. Your willingness to jump into this project with both feet, along with your feedback and guidance throughout this process has been invaluable.

To Loyola’s Office of Research Services, and especially Andrew Ellis, thank you for allowing me the opportunity to explore another aspect of the research process through the IRB. Also, thank you to The Graduate School for sponsoring the Dissertation Award that provided partial funding for this project. To the faculty and staff of the Psychology Department, thank you for always reminding me that knowledge is power, hard work is important, and the key to a great party is the food and drink selection.

To my cohort, Chad, Darian, Josh, Kari, Linas, Liza, Michelle, and Milan, I have never felt such an immediate, incredible bond with another group of humans. The time we have shared…Who am I kidding? The fun does not end here – join me at the afterparty. To the graduate students that came before me, thank you for your words of wisdom and free textbooks. To those that will come after me, I’m always an email, text, or phone call away.
To my family and friends, I am so thankful that you stuck by me even after the countless ignored texts and phone calls, and weekend plans cut short by work. Thank you, mom, for remaining my sounding board and keeping me grounded. I promise to continue to make you proud. Dad, thank you for always being the first to volunteer as my “test subject”. To my brother, Bill, you are one of the smartest people I know. I will always be learning from you. Thanks to my niece, Ruby, for being the cutest part of graduate school. To my best friend, Sam, you are a profoundly caring and reliable person. Your understanding throughout this time is unmatched. Let’s finally grab that beer this weekend. Dr. Robert Cole (Bob), my academic confidant and amazing friend, thank you for the long venting sessions and statistics help.

Finally, I owe it all to the person whose support has remained steadfast through the laughter and the tears. Cayla, you are truly my rock – I could never have done this without you. Cheers to our next adventure!
TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
LIST OF TABLES viii
LIST OF FIGURES x

ABSTRACT xi

CHAPTER ONE: THE PROBLEM
How Did We Get Here? Understanding Racism 1
Working Toward A Solution 5

CHAPTER TWO: CONFRONTATION
Outcomes of Confronting: Perpetrators’ Prejudice Reduction 7
Outcomes of Confronting: Social Backlash for the Confronter 9
The Role of White Allies 10
Barriers to Confronting 11
Exploring the Role of Social Bonds in the Outcomes of Confrontation 13

CHAPTER THREE: HARNESSING THE POWER OF SOCIAL BONDS 15
Overview of Social Bonds 15
Threats to Social Bonds 17
Outcomes of Confrontation in Social Bonds: Prejudice Reduction 18
Emotional Reactions to Confrontation in Social Bonds 20
Outcomes of Confrontation in Social Bonds: Social Backlash 21
The Unique Influence of Trust 23

CHAPTER FOUR: STUDY DESIGN
Hypotheses 28
Study 1 31
Study 1 Method 33
Study 1 Results 37
Study 1 Discussion 45
Study 2 48
Study 2 Method 49
Study 2 Results 52
Study 2 Discussion 65

CHAPTER FIVE: GENERAL DISCUSSION
Limitations and Future Directions 73
Conclusion 74
APPENDIX A: ONLINE RECRUITMENT TEXT 76
APPENDIX B: STUDY 1 INFORMED CONSENT 78
APPENDIX C: INTRODUCTION COVER STORY AND AVATAR SELECTION 80
APPENDIX D: AFFILIATIVE MOTIVATION PROMPT 82
APPENDIX E: AFFILIATIVE MOTIVATION CHECKS 84
APPENDIX F: TRUST MEASURES 86
APPENDIX G: PHOTO-INFERENCE TASK 88
APPENDIX H: CONFRONTATION FEEDBACK 92
APPENDIX I: NESELF & NEGOTHER QUESTIONNAIRE 94
APPENDIX J: DEMOGRAPHICS 96
APPENDIX K: HYPOTHESIS GUESS & PERFORMANCE EVALUATION 98
APPENDIX L: STUDY 1 DEBRIEFING FORM 100
APPENDIX M: STUDY 1 DATA SCREENING DETAILS 102
APPENDIX N: STUDY 1 SUPPLEMENTARY RESULTS 104
APPENDIX O: STUDY 2 INFORMED CONSENT 116
APPENDIX P: STUDY 2 INTRODUCTION ADDITIONS 118
APPENDIX Q: STUDY 2 CONFRONTATION FEEDBACK ADDITIONS 120
APPENDIX R: STUDY 2 DEBRIEFING FORM 122
APPENDIX S: LIKING MEASURE 124
APPENDIX T: ATTITUDES MEASURE 126
APPENDIX U: STEREOTYPE ENDORSEMENT MEASURE 128
APPENDIX V: INTETIONS TO CONTROL FUTURE BIAS MEASURE 130
APPENDIX W: STUDY 2 DATA SCREENING DETAILS 132
APPENDIX X: STUDY 2 SUPPLEMENTARY RESULTS 134

REFERENCE LIST 150

VITA 163
LIST OF TABLES

Table 1. Bivariate correlation matrix detailing the associations between variables of interest. 40

Table 2. ANOVA results detailing the effects of affiliative motivation, confrontation type, and trust on negative self- and other-directed emotions. 42

Table 3. PROCESS Model 2 results for Hypothesis 1a depicting the interaction between affiliative motivation and trust measured before and after a confrontation on negative self- and other-directed emotions. 44

Table 4. Bivariate correlation matrix detailing the associations between variables of interest. 55

Table 5. ANOVA results detailing the effects of affiliative motivation and confrontation type on all variables of interest. 58

Table 6. PROCESS Model 8 results depicting the index of moderated mediation for the full model with affiliative motivation serving as the independent variable, negative self- and other-directed emotions acting as mediators, and backlash and prejudice reduction outcomes as dependent variables. 59

Table 7. PROCESS Model 8 results depicting the interaction of affiliative motivation and trust on intentions to control future bias after a racial confrontation. 64

Table 8. Details of the data screening process for participant race (Step 1). 103

Table 9. Details of the data screening for adhering to the procedure (Step 3). 103

Table 10. Details of the supplemental data screening process for participant race (Step 1). 107

Table 11. Details of the supplemental data screening process for the affiliative motivation manipulation check (Step 3). 107

Table 12. Details of the supplemental data screening process for adhering to the procedure (Step 4). 108

Table 13. Supplemental bivariate correlation matrix detailing the associations between variables of interest. 109
Table 14. Supplemental ANOVA results detailing the effects of affiliative motivation, confrontation type, and trust time point on negative self- and other-directed emotions.

Table 15. Supplemental PROCESS Model 2 results for Hypothesis 1a depicting the interaction between affiliative motivation and trust measured before and after a confrontation on negative self- and other-directed emotions.

Table 16. Details of the data screening process for participant race (Step 1).

Table 17. Details of the data screening process for adhering to the procedure (Step 3).

Table 18. Details of the supplemental data screening process for participant race (Step 1).

Table 19. Details of the supplemental data screening process for the affiliative motivation manipulation check (Step 3).

Table 20. Details of the supplemental data screening process for adhering to the procedure (Step 4).

Table 21. Supplemental bivariate correlation matrix detailing the associations between variables of interest.

Table 22. Supplemental ANOVA results detailing the effects of affiliative motivation and confrontation type on all variables of interest.

Table 23. PROCESS Model 8 supplemental results depicting the index of moderated mediation for the full model with affiliative motivation serving as the independent variable, negative self- and other-directed emotions acting as mediators, and backlash and prejudice reduction outcomes as dependent variables.

Table 24. PROCESS Model 8 supplemental results depicting the interaction of affiliative motivation and trust on intentions to control future bias after a racial confrontation.
LIST OF FIGURES

Figure 1. Moderated mediation of the indirect effects of NegSelf and NegOther on the association between affiliative motivation and prejudice reduction and backlash by trust for racial (v. rude) confrontations. Path letters correspond to hypotheses 1, 3, and 4. 31

Figure 2. Details of the data screening process for Study 1. 38

Figure 3. Plot depicting trust measured after the confrontation (collapsed across condition) moderating the association between affiliative motivation and negative other-directed emotions (hypothesis 1b). 45

Figure 4. Details of the data screening process for Study 2. 53

Figure 5. Plot depicting trust moderating the association between affiliative motivation and intentions to control future bias after a racial confrontation. 64

Figure 6. Details of the supplemental data screening process for Study 1. 106

Figure 7. Supplemental plot depicting trust measured after the confrontation (collapsed across condition) moderating the association between affiliative motivation and negative other-directed emotions (hypothesis 1b). 114

Figure 8. Details of the supplemental data screening process for Study 2. 136

Figure 9. Plot depicting the supplemental results of trust moderating the association between affiliative motivation and intentions to control future bias after a racial confrontation. 148
ABSTRACT

Anti-Black racism persists in the United States with harmful consequences for Black people. White people are able to disrupt the racial status quo and propel the conversation about racial justice forward by confronting racism. Confronting perpetrators of racism can reduce prejudice, yet people hesitate to confront because they fear social backlash, even from those with whom they share a social bond. Two online studies asked participants to complete a task eliciting stereotypical responding while being observed by a supposed interaction partner with whom they shared either a high or low desire to get along. Participants were confronted by their interaction partner for racism or in a rude way. I tested whether a perpetrator’s trust in a confronter would influence the strength of the association between their motivation to form a social bond and the outcomes of a racial confrontation (i.e., negative emotionality, backlash, and prejudice reduction). As predicted, results showed that relative to participants with high affiliative motivation, trusting a confronter with whom one has low affiliative motivation decreased negative confronter-directed emotions (Study 1, \( n = 538 \)) and backlash, and increased intentions to control future bias (Study 2, \( n = 869 \)). Contrary to predictions, negative self-directed emotions following a confrontation for racism increased negative stereotype endorsement and attitudes toward Black people. These results indicate that harnessing the power of trust might mitigate backlash for White ally confronters who are strangers, while promoting a perpetrator’s behavior change. The effectiveness of online manipulations of affiliative motivation as well as the impact of this moment in history are discussed in terms of their influence on the results of these studies.
CHAPTER ONE

THE PROBLEM

On May 25, 2020, a Minneapolis police officer, Derek Chauvin, was captured on video kneeling on George Floyd’s neck. Floyd, a Black man whose alleged crime was passing a counterfeit $20 bill at a local corner store, struggled to breath for just over 8 minutes and 46 seconds until his death. The murder of George Floyd sparked renewed civil uprisings throughout the United States on behalf of all Black people killed at the hands of police brutality and anti-Black racism (Weine et al., 2020). Twenty years after the turn of the 21st century, the United States of America was forced to battle concurrent pandemics. Alongside the outbreak of a global illness caused by the novel coronavirus, the nation once again reckoned with widespread racism wreaking havoc on the lives of Black people as it had for centuries (Barbot, 2020). Found within every system, community, and household, anti-Black racism has taken countless lives 155 years after the abolishment of slavery and over 50 years since the adoption of landmark civil rights laws. Perhaps most visibly, bias inherent in the criminal justice system has taken the lives of Breonna Taylor as she slept in her bedroom, Attatiana Jefferson as she played video games with her nephew, Botham Jean as he ate ice cream on his sofa, and Tamir Rice as he played in the park. Stephon Clark, Phillando Castille, Michelle Cousseaux, Alton Sterling, Eric Garner, Sandra Bland, Freddie Gray, Michael Brown; the list of names of Black people killed by those who are tasked to protect and serve cannot be contained in a single paragraph, page, or dissertation. The driving factor behind these
deaths rests with the anti-Black bias held by individuals and revealed through their personal thoughts, words, and actions (Takagi, 1974).

Motivated by their determination to reduce anti-Blackness, some White people have participated in the ongoing conversation about racial injustice through protests and social uprisings such as the Black Lives Matter movement. However, White people who remain in denial or even in support of anti-Blackness present a barrier for those attempting to move the conversation about racial justice forward (Holt, 2018). Thus, White people are at a crossroads at which they must decide to either challenge anti-Black racism to reduce its occurrence or remain silent and allow its continuation. For those choosing to confront, starting with perpetrators with whom they share a social bond is a logical place to begin.

**How Did We Get Here? Understanding Racism**

Anti-Black racism, or the expression of prejudice targeting Black people, has taken on many forms over centuries. It has persisted from the Colonial era through the period of Reconstruction, Jim Crow segregation, and most recently, Civil Rights legislation and other federal mandates deeming discrimination against Black people illegal (Dovidio et al., 2017). Psychologists have studied anti-Black racism through its morphosis from the old-fashioned version (i.e., the overt expression of negative evaluations of another racial group), to aversive (i.e., the denial of negative feelings and general avoidance of racial minorities) and modern forms (i.e., the denial of the existence of racial discrimination altogether; Gawronski et al., 2010).

At the center of racism is oppression, the notion that some groups (e.g., White people) are dominant over other groups (e.g., Black people), and the dominant White groups use their
conferred authority to marginalize groups of color (David et al., 2019). One way to identify anti-Black racism is through systems allowing dominant institutions and individuals power over subjectively inferior others (Jones, 1972) by enforcing the laws, customs, and practices leading to the mistreatment of racial minorities (Deutsch, 2006). From healthcare to education, housing to politics, anti-Black racism can be identified in most systems with those in power often upholding the tradition. Referred to as institutional racism, this practice signals and amplifies the power and influence of White people (Kelly & Varghese, 2018) in the United States.

Privilege, as it has been granted to certain people based solely on their White racial group membership regardless of whether they desire or are aware of it, allows for economic, social, and political advantages over people of color (David et al., 2019). White people have been enculturated to believe in their superiority because to face their privilege means forgoing the myth of meritocracy, or their stance that hard work is to credit for their successes (Costa-Lopes et al., 2013). When left unchecked, the unconscious beliefs of primacy held by White people affect their daily behavior and often lead to anti-Black prejudice (Rudman & Ashmore, 2007) and the interpersonal racism seen in the United States today.

Institutional and interpersonal racism work hand in hand, with institutional racism operating at a macro-level and interpersonal racism operating at a more micro-level. While institutional racism is found in policies and procedures that favor one racial group or fail to create equity, interpersonal racism involves more powerful and privileged White individuals engaging in biased attitudes, thoughts, and behaviors toward individual people of color (Jones, 1972; David et al., 2019). Interpersonal racism is often more easily seen and acknowledged compared to institutional racism (Ture & Hamilton, 1968), in part because White people created
the systems, making them seem legitimate and correct. For example, police officers are more likely to use deadly force on a Black (versus White) person (Trawalter et al., 2019). Law enforcement officers shoot unarmed Black victims more often than their White counterparts (DeGue et al., 2016). Stereotypes of Black people as violent criminals (Hurwitz & Peffley, 1997) and the resulting interpersonal discrimination better explain these findings than the actual rates of violent crime committed by Black people (Mesic et al., 2018). Additionally, White landlords, bankers, and realtors’ interpersonal discriminatory practices in offering housing options contribute to homelessness experienced by Black people (Feagin, 1999). White healthcare providers have also been found to hold racist beliefs and emotions regarding their racial minority group patients (Paradies et al., 2014), and continue to assume that Black people feel less pain than White people (Trawalter & Hoffman, 2015). Thus, the institutional racism divulged through law enforcement, housing, and healthcare services is upheld and reinforced by the interpersonal racism experienced during a traffic stop, or in a bank or doctor’s office.

However, experiences with anti-Black racism are not confined to those abstract datasets regarding health statistics or news stories about police brutality. Interpersonal racism occurs daily, in any context. Black people report experiencing “everyday racism”, which encompasses direct racist comments and subtle biased behaviors perpetrated by White individuals in their environment (Swim et al., 2003), such as verbal expressions of prejudice, staring, bad service, and awkward or rude exchanges. Everyday racism is not context-limited in that it occurs frequently during social interactions in a variety of environments including the classroom, home, and workplace. The frequency of this interpersonal anti-Black racism has led people who may be categorized as Black to anticipate the bias as a function of survival or “a fact of life” (Krieger &
Sidney, 1996), often resulting in psychological distress in the form of depression, anger, and anxiety (Stangor et al., 2003). This means that White perpetrators of racism are not confined to a particular context either, allowing their bias to be challenged in the classroom, workplace, or at home at the dinner table.

**Working Toward a Solution**

Since the 2016 presidential election, norms have shifted to permit more expressions of prejudice (Crandall et al., 2018). Negative views toward racial or ethnic minorities are common and acceptable (Horowitz et al., 2019), demonstrating that anti-Black racism remains alive within the United States (Huber, 2016). One reason anti-Black racism persists is that White people are socialized in and possess a racial identity tied to White supremacy (Carter et al., 2004). Subsequently, White people internalize and embrace status-legitimizing beliefs and practices, which suggest that their position of power and privilege within society has been earned with hard work (Jost & Banaji, 1994; Jost et al., 2004; Jost & Hunyady, 2005). Moreover, claims of racism by Black people are often seen as a threat to the social hierarchy as suggestions of inequality prove the existing hierarchy to be illegitimate (Wellman et al., 2016). With that, individuals with White racial group membership operate in a world in which the deck has been deliberately stacked against those with Black racial group membership and behave in a way, whether deliberate or not, that maintains their power and privilege and hurts Black people regardless of intent.

This work tests interpersonal solutions to anti-Black racism. One such solution is to identify the conditions under which White people might be more receptive to confrontational messages related to reducing the harm of racism for Black people, starting within their own
relationships. The present research identifies ways to maximize the potential for White people to engage in the ongoing conversations and strategies proposed to end White supremacy. Talking about anti-Black racism with White perpetrators within one’s inner circle can be seen as the first step in dismantling this persistent social problem.
CHAPTER TWO

CONFRONTATION

Most people think of confrontation as possessing an argumentative connotation that seems counterintuitive to effective interpersonal conversations aimed at reducing anti-Black racism. In comparison, the field of social psychology defines confrontation as a strategy for responding to prejudice, encompassing any signal of opposition to the biased attitude or behavior of a perpetrator (Barreto & Ellemers, 2015; Shelton et al., 2006). Confrontations can come in a variety of verbal declarations (e.g., expressing disapproval, responding with sarcasm, questioning the perpetrator, providing a biased label for the behavior, reporting the incident to a third party) or non-verbal behaviors (e.g., emoting surprise or disgust, remaining silent, leaving the setting; Dickter & Newton, 2013; Shelton & Stewart, 2004; Shelton et al., 2006; Woodzicka & LaFrance, 2001; Swim & Hyers, 1999). Thus, confrontations of prejudice do not necessitate the hostility many people think of when considering it as a bias intervention. In fact, confrontations often result in beneficial outcomes.

**Outcomes of Confronting: Perpetrators’ Prejudice Reduction**

Confronting perpetrators of bias has been shown to reduce prejudice (Czopp et al., 2006; Czopp & Monteith, 2003; Gulker et al., 2013). Prejudice reduction has been operationalized in terms of attitude change, or less application and endorsement of stereotypes. For example, in three studies investigating the outcomes of racial confrontations conducted by Czopp and colleagues (2006), naïve participants completed a photo-inference task designed to evoke
stereotypical assumptions about Black targets presented in photographs with ambiguous captions (e.g., “This person can be found behind bars” elicited the stereotypical response, “criminal”, rather than “bartender”). Participants were then confronted for their biased responses by a confederate. Confronted participants were less likely to provide stereotypic responses on a subsequent photo-inference task than participants who were not confronted (Study 2), denoting an example of a change in stereotype application. Moreover, following a confrontation, participants reported less prejudiced attitudes (Study 3). Confrontations for racism and sexism decrease the endorsement of negative stereotypes toward multiple stigmatized groups across identity domains regardless of the focus of the allegation of bias. Chaney and colleagues (2020) found that compared to participants who did not receive a confrontation, those confronted for using negative Black stereotypes, subsequently used fewer negative Latino stereotypes. Likewise, participants confronted for using female gender role stereotypes subsequently used fewer Black and Latino stereotypes.

Prejudice reduction has also been measured as a change in behavior and behavioral intentions. For example, following a confrontation for sexism, Mallett and Wagner (2011) found that men engaged in more compensatory behaviors, such as smiling or offering an apology, and were better able to detect the use of sexist language in a subsequent task. More recently, researchers have begun to measure concern about future bias to assess prejudice reduction. In a series of studies, Parker and colleagues (2018) explored the effectiveness of evidence-based confrontations to regulate future gender bias. The authors found that when participants were confronted for sexism with concrete evidence of their personal bias, they scored higher on a measure of intentions to control future bias compared to participants confronted without evidence.
One mechanism through which confrontations of bias work to reduce prejudice is negative self-directed affect (e.g., guilt, self-criticism). In the aforementioned study, Czopp and colleagues (2006) demonstrated that as perpetrators felt more negative self-directed emotions following a confrontation for racism, their stereotypic responses in the subsequent photo-inference task decreased as did their negative evaluations of the confronter (see replication Hypothesis 2). The presence of prolonged guilt and rumination following an interpersonal confrontation contributes to the endurance of behavioral change for at least one week (Chaney & Sanchez, 2018). However, some perpetrators of bias externalize their negative feelings after a confrontation rather than directing negative affect inward (Monteith et al., 1993).

**Outcomes of Confronting: Social Backlash for the Confronter**

Allegations of racial bias via confrontation can simultaneously engender negative other-directed affect (e.g., anger, irritation) aimed at the confronter (Czopp et al., 2006). Because White people seek to be liked in interracial communications and fear being seen as racist (Bergsieker et al., 2010), they often fail to acknowledge that interpersonal racism exists. Ignoring allegations of racism is often due to the perceived personal threat of being called racist, as White people on the receiving end of a racist accusation are motivated to maintain their own positive self-image (Unzueta & Lowery, 2008). In response to this threat to the self, feelings of anger and irritation arise often leading to decreased prejudice reduction and increased social consequences or “backlash” (Monteith et al., 1993), the penalties incurred by the confronter for alleging racism (see replication Hypothesis 2). The potential for social consequences often stymies confrontation. Though targets of discrimination (i.e., individuals who hold a historically oppressed identity) often imagine that they would confront a perpetrator’s bias, few actually do so (Woodzicka & LaFrance, 2001). After reading a vignette in which a male interviewer asked
sexist questions, Shelton and Stewart (2004) observed that women reported an overwhelming intention to confront the perpetrator regardless of the magnitude of the associated social consequences. However, a second study conducted in a face-to-face manner revealed that in a high (i.e., sexist questions are asked to inform the interviewer’s hiring decision) versus low (i.e., sexist questions are asked to peak the interviewer’s curiosity) social cost scenario, women were less likely to confront individuals who behave in a prejudiced manner. Targets of discrimination refrain from confronting biased perpetrators due to the anticipation of amplified backlash compared to individuals who do not possess the targeted identity (Swim & Hyers, 1999; Fitzgerald et al., 1995). There is evidence to support the idea that targets of prejudice experience more social consequences for confronting than non-targets. Witnesses of racial discrimination evaluate Black people who confront racism more negatively than White people who confront racism (Rasinski & Czopp, 2010; Schultz & Maddox, 2013). Thus, Black people are often reluctant to make attributions of discrimination for unfair treatment (Kaiser & Miller, 2001) or acknowledge bias altogether (Kaiser & Miller, 2003) because of the increased likelihood of being labeled a complainer, troublemaker, hypersensitive, or irritating.

**The Role of White Allies**

Allies are individuals who support a historically disadvantaged group by acting on the need for progress toward equal rights (Washington & Evans, 1991). Although confrontations of racism delivered by Black and White people have been shown to be equally effective in reducing biased behaviors and attitudes (Czopp et al., 2006), there is also evidence to suggest that when allies confront, they are seen as more serious and legitimate and are positively evaluated (Drury & Kaiser, 2014). In terms of racism, confrontations by White allies are often met with greater acceptance (Gulker et al., 2013), as they are more likely to be deemed accurate, true, and in need
of future attention. Compared to those delivered by targets, allied confrontations are often rated as more persuasive (Rasinski & Czopp, 2010), and observers feel increased liking and respect for White people who confront offensive racist remarks within their ingroup compared to other White people who do not confront racist remarks (Dickter et al., 2012). The differences in the social consequences of confronting prejudice can be explained by the intergroup sensitivity effect, whereby group-based criticisms delivered by outgroup members are viewed as less constructive than those coming from an ingroup member (Hornsey et al., 2002). Moreover, people respond more defensively to outgroup members’ criticisms (Hornsey & Imani, 2004). Although it has not yet been tested, the intergroup sensitivity effect may allow White people to confront ingroup members for racism while dodging backlash and reducing prejudice.

**Barriers to Confronting**

Aside from the concerns about social sanctions, numerous other barriers exist. These barriers are often self-imposed due to a lack of understanding of bias, preventing White people from challenging racism. Ashburn-Nardo and colleagues (2008) outline the hurdles involved in confronting in the Confronting Prejudiced Responses (CPR) model. The CPR model can explain White silence and inaction in the face of anti-Black racism. The model’s five steps, based on the bystander intervention work by Latané and Darley (1969), describe what a White ally needs to grapple with to go from observing or experiencing a discriminatory situation to actually confronting the perpetrator.

The first step of the CPR model requires that one detect discrimination and label it as such. While this step might seem straightforward, it can be challenging for White people when the discriminatory event does not match prototypical definitions (Martí et al., 2000), such as denying public services or providing differential treatment based on certain identities (Hebl et
al., 2002). For example, White perpetrators often construct alternative explanations for racially biased behaviors to uphold their egalitarian appearance and values (Dovidio et al., 2002), making it even more difficult for a White ally to recognize racism. Second, people must interpret the discrimination as an emergency, which can be difficult since most discrimination occurs without physical harm to targets and lacks a signal of other types of harm (Fehr et al., 1979). Third, the individual must take responsibility for confronting the incident. The majority of White people may not feel it is their place to confront when a target is present (Crosby et al., 2008). Fourth, the observer must have the appropriate strategies to confront. Due to the frequency of Black Americans’ experiences with racism, they are more likely than their White counterparts to teach their children about prejudice as well as how to respond when one is the target of discrimination (Hughes et al., 2006; Lesane-Brown, 2006). Therefore, it is unsurprising that White people, who do not often receive an education about challenging prejudice and lack confidence about how to do so effectively (Martinez et al., 2017), are unlikely to confront bias on behalf of racial minorities (Kawakami et al., 2009).

Finally, if the aforementioned conditions have been met and all of the steps have been undertaken, Ashburn-Nardo and colleagues (2008) posit that the observer will confront the discriminatory situation if the benefits outweigh the costs. However, even if the observer makes it to this final step, the perceived social consequences of confronting could still outweigh the perceived benefits. Additionally, the opportunity to confront racism perpetrated by their racial ingroup members is not rare for White people, making the lack of confrontation even more troublesome. Dickter and Newton (2013), for instance, found that after keeping a log of prejudice, White undergraduate students heard almost nine direct or indirect racist comments by other majority group members over the course of one week. White students, however, were
unlikely to confront the people who made these comments, even though they were frequently family, friends, or acquaintances. The evidence in this case suggested that bypassing the opportunity to confront was partly driven by the fear of damaging the relationship between the confronter and the perpetrator of discrimination, a fear that can ultimately hinder opportunities to challenge bias for many people (Good et al., 2012).

**Exploring the Role of Social Bonds in the Outcomes of Confrontation**

In summary, White allies are in a unique position to challenge perpetrators of discrimination as their confrontations are not only effective in reducing biased attitudes (Czopp et al., 2006), but White people are likely to incur fewer interpersonal consequences (i.e., backlash) compared to Black confronters (Drury & Kasier, 2014). However, they anticipate negative outcomes, such as social rejection (Swim & Hyers, 1999; Woodzicka & LaFrance, 2001), which could be remedied by exploring the outcomes of confrontations within dyads who share a social bond. Research to date has found that the social bond shared between the perpetrator and confronter of bias plays a role in the likelihood of confrontation. For example, targets of discrimination are more likely to engage in confrontation if the perpetrator is someone they personally know or of equal status, and less likely if the perpetrator is a stranger (Ayres et al., 2009). Moreover, when individuals feel supported and believe that confronting a perpetrator will elicit change, such as in interactions with equal status friends, they are more likely to confront (Buchanan et al., 2007). Previous research has only explored the likelihood of confronting within socially bonded pairs and the outcomes of confrontations between strangers. The literature has yet to identify the outcomes of confrontations between people who share a social bond, and the aforementioned studies provide reason to believe that the outcomes of racial confrontations, specifically prejudice reduction and backlash, might differ depending on the
relationship between the perpetrator of discrimination and the confronter. Perhaps the existence of the motivation to get along with a confronter can reach beyond increasing the likelihood of confronting bias (Ayres et al., 2009; Buchanan et al., 2007), to actually effecting the outcomes of the confrontation. With that notion in mind, an opportunity exists for White allies to challenge discrimination in their daily lives with people with whom they are closest, while simultaneously reducing prejudice and avoiding social backlash.
CHAPTER THREE

HARNESSING THE POWER OF SOCIAL BONDS

The drive to avoid social consequences may be one factor that undermines confrontation of anti-Black racism. However, leveraging social bonds with White perpetrators may empower White allies to confront bias with less fear of backlash. Research has documented White allies’ fear of losing social connections (Good et al., 2012) during confrontations of bias between strangers. The literature on close relationships (see Murray et al., 2006; Murray et al., 1996) provides some clues that outcomes of confrontation between socially bonded pairs may differ compared to those lacking any relational bond.

Overview of Social Bonds

Human beings are social by nature. Our propensity to bond with others bolsters our survival as a species (Taylor, 2012) and allows us to flourish across the lifespan (Taylor, 2009). In order to solidify social bonds, humans utilize self-expansion or the act of including others in one’s sense of self (Aron & Aron, 1993). As people increasingly include another in their sense of self, they feel closer to that person, resulting in the establishment of a bond. Maslow (1967) proposed that humans must satisfy basic motives like hunger and safety. Aron and colleagues (2008) argue that forming and maintaining social bonds might allow for more efficient satisfaction of those basic needs. This underscores the importance of forming and maintaining social bonds, despite threats to those relationships. Although much of the work on social bonds is derived from intimate relationships, all social bonds, regardless of their history or intensity,
include key components that allow for comparisons between them. For example, shared reality is an essential ingredient to any social relationship (Hardin & Higgins, 1996). People actively seek reassurance and certainty about their beliefs and values (Echterhoff et al., 2009), and developing a mutual understanding (Sinclair & Skorinko, 2018), or shared reality (Hardin & Higgins, 1996; Hardin & Conley, 2001), with others helps to legitimize those beliefs and values. Ultimately, the creation of shared reality with another satisfies the fundamental need to form and preserve social relationships, and understand the world (Echterhoff et al., 2009).

To achieve shared reality, people must think deeply about another person’s point of view or take their perspective (Epley & Caruso, 2008). Even in those cases when the target with whom one attempts to share reality is unfamiliar, perspective taking allows for the initiation of self-other overlap (Davis et al., 1996). Individuals including another in the self will ascribe more positive attributes of the other person into their own self-descriptions, facilitating social coordination and forming a relationship (Galinsky et al., 2005). Thus, previous research focused on close relationships in which shared reality, perspective-taking, and self-other overlap are abundant can provide insight into the ways in which more tenuous social bonds might operate.

In general, forming and maintaining social bonds (Blieszner, 2014; Cable et al., 2013) makes people feel good. That positive affect is often transferred onto bonded partners as a reflection of the relationship formation and maintenance process (Aron et al., 2008). The positive affect people feel when developing bonds leads to positive illusions about their partner’s virtues (e.g., intelligence, understanding, wit). For instance, when thinking of their friends, people are more likely to provide positively valanced virtue evaluations compared to when thinking about the average other, or a stranger (Martz et al., 1998). Positive illusions enhance relationship satisfaction (Murray et al., 1996), and may also be adaptive during conflict or threat.
Threats to Social Bonds

All relationships inevitably face threats, particularly those that come in the form of interpersonal conflict. Still, in the face of relationship conflict, people will actively construct idealized impressions of their partners (Murray, 1999), not in spite of their faults but because of their partner’s imperfections (Murray & Holmes, 1993). Taylor and Brown (1988) suggest that positive illusions of close others promote the ability to remain caring and content with one’s relationship partner, even when faced with their negative feedback. Without positive illusions of close others, relationship satisfaction suffers when a partner experiences a relational threat in the form of unpleasant interpersonal events, like expressions of criticism (Huston & Vangelisti, 1991). In this way, the positive illusion that exists in the context of a social bond might protect one’s sense of satisfaction with and caring for their partner despite receiving negative feedback about their bias.

In a practical sense, sharing reality and the ability to interpret cues regarding how those close to us think and feel enables us to determine how we think and feel. For example, shared reality might clue us into which movies to watch or political candidate to select (Hardin & Conley, 2001; Hardin & Higgins, 1996), or even how to evaluate outgroup members (Sinclair et al., 2005). As such, the motivation to maintain a shared reality and get along with another, called affiliative motivation, might foster avoidance of relational threats and conflict by allowing a perpetrator of racism to align their attitudes and behavior with the egalitarian values of a their socially bonded confronter.

We know that confrontations of bias between two strangers may result in prejudice reduction (Czopp & Monteith, 2003; Czopp et al., 2006), but often produce social backlash (Eliezer & Major, 2012; Kaiser & Major, 2006). This research tested whether confrontations that
Outcomes of Confrontation in Social Bonds: Prejudice Reduction

One avenue toward achieving shared reality is to “tune” one’s beliefs toward another’s (Skorinko & Sinclair, 2018). The process of social tuning or aligning one’s attitudes with those of an interaction partner, is driven by affiliative motivation, or a desire to form or maintain social bonds (Sinclair et al., 2005). In much of this research, affiliative motivation was achieved by varying two aspects of an impending social interaction. Participants induced to feel high affiliative motivation were told that they shared the same birthday as their partner and that the two individuals would interact for an extended period of time (i.e., 30 minutes). Those induced to feel low affiliative motivation were informed that their birthdays differed from their partner and that they would be interacting for only 5 minutes. When affiliative motivation was high, individuals spontaneously adopted the attitudes of individuals with whom they interacted, and when affiliative motivation was low, individuals adjusted their views away from those of an interaction partner as a strategy to distance themselves (i.e., anti-tuning). Using this methodology, Sinclair and colleagues (2005) found that Black students portrayed themselves as conforming more or less to stereotypes of Black people depending on the presumed views of their partner. Black students with high affiliative motivation for an interaction partner with stereotypical views of their group engaged in self-stereotyping. Black students with low affiliative motivation for a similarly biased individual formed counter-stereotypic views of themselves.
Inducing affiliative motivation is a well vetted and particularly useful technique when the goal is to randomly assign people to feel a connection to another person in a lab setting. Similar procedures have also been used in an online chat environment with the primary manipulation being the friendliness of the ostensible interaction partner, whereby participants in the high (versus low) affiliative motivation condition interacted with a friendly (versus unfriendly) partner (Huntsinger et al., 2016). Regardless of the setting or details of the manipulation, high (versus low) affiliative motivation is found to move individuals toward aligning their beliefs with those apparently held by the other half of the socially bonded pair.

Social tuning as the result of affiliative motivation is a powerful tool by which negative intergroup attitudes can be reduced. If the perpetrator is motivated to affiliate with the confronter, then the perpetrator may socially tune their racial attitudes toward the expressed egalitarian attitudes of the confronter. Along these lines, Sinclair and colleagues (2005) found that participants who were assisted through a study by an experimenter wearing a shirt with an anti-racist logo (i.e., Eracism) tuned their implicit racial bias toward the egalitarian experimenter and exhibited less automatic racial prejudice compared to those who were assisted by an experimenter wearing a plain shirt. The inherent desire to maintain social bonds, or affiliative motivation, within novel relationships supports the hypothesis that confrontations within socially bonded pairs may reduce prejudice. In other words, if a person with racist attitudes (i.e., a perpetrator) is motivated to get along with a confronter who expresses egalitarianism, then tuning their attitudes toward the confronter should reduce bias.

Another path toward prejudice reduction might be through something I will call a relational sensitivity effect. As previously noted, the intergroup sensitivity effect proposed that criticisms delivered by outgroup members were met with more defensiveness than those
delivered by an ingroup member (Hornsey et al., 2002; Hornsey & Imani, 2004). The intergroup sensitivity effect is based on the notion that individuals derive part of their self-concept from salient group memberships (Tajfel, 1978; Tajfel & Turner, 1979; Turner, 1999). It follows that we may observe a similar relational sensitivity effect, whereby outgroup members are like strangers and ingroup members are like those in our social circle with whom we share a bond. Threats delivered by ingroup members can serve a positive function, such as innovation, creativity, and flexibility in decision making (Nemeth & Owens, 1996). Thus, if there is a relational sensitivity effect, then perpetrators of racism may react more defensively when confronted by a stranger than when confronted by someone with whom they are bonded. Perpetrators who are confronted by a bonded other might respond with flexibility in their thinking and be more likely to change their biased attitudes and behavior.

**Emotional Reactions to Confrontation in Social Bonds**

Emotions play an essential role in the prejudice reduction process for White perpetrators of anti-Black bias (Czopp et al., 2006). One type of emotional reaction to confrontations of bias is negative self-directed affect, which includes feelings of compunction like self-criticism or guilt (Monteith et al., 1993). Guilt is a negative self-directed emotion that can engender attitude change. According to Baumeister and colleagues (1994), guilt exists as a product of the acknowledgement and acceptance that a transgressor has inflicted harm, loss, or distress upon another within a communal relationship. Guilt reactions are more common and stronger within close relationships compared to nonexistent relationships. The researchers note that as the commonality between two people approaches zero, so too does the possibility for guilt. It follows that guilt might be amplified after being confronted by a bonded other compared to that found
between total strangers. These increased feelings of guilt after a confrontation between two people motivated to get along could result in increased prejudice reduction (see Hypothesis 3).

Although guilt has a negative connotation, it actually enhances relationships by eliciting symbolic affirmation of caring and commitment (Baumeister et al., 1994). Feelings of guilt remind the perpetrator that they care about and are committed to the relationship and their relationship partner, triggering the guilty party to engage in reparative action to mend the damage their actions have caused to the relationship. Feeling threatened by guilty emotions within a social bond should enhance relationship promotion and dependence in pairs who are motivated to form or maintain a relationship (Baumeister et al., 1994). In contrast, interactions lacking a social bond should be less likely to result in the negative self-directed emotion of guilt and may elicit increased feelings of global discomfort (i.e., negative other-directed affect) projected from the perpetrator toward the confronter. This negative other-directed affect, or the perpetrator feeling bothered or threatened (Monteith et al., 1993) has commonly been associated with social backlash targeting confronters (Czopp et al., 2006; see Hypothesis 3).

**Outcomes of Confrontation in Social Bonds: Social Backlash**

The theory of emotional capital posits that every relationship possesses an emotional “bank account” into which positive, shared experiences (e.g., compliments, smiles, laughter, engagement in fun activities) are deposited (Driver & Gottman, 2004). Partners in established relationships exhibit high emotional capital and are less likely than individuals lacking a social bond to show declines in positive feelings or enact negative behaviors toward one another (Feeney & Lemay, 2012) after experiencing damaging or hurtful behaviors (e.g., criticism). Thus, socially bonded pairs who have made deposits into their emotional bank accounts might be less likely than those without a social bond who have made no such contributions to react
negatively toward a confronter of bias. Likewise, individuals high (versus low) in affiliative motivation engage in behaviors that are guided by relationship goals and generally act less aggressively (Hill & Werner, 2006; Delveaux & Daniels, 2000). In my prior unpublished work (Patrianakos, 2018), high (versus low) affiliative motivation protected confronters from backlash after an aggressive (versus polite) confrontation for racism. Additionally, replication of the same study procedures using real friendship pairs rather than an affiliative motivation procedure produced identical results. Results of these studies provide further evidence that novel social bonds created using an affiliative motivation to share reality operate in much the same way as established relationship (e.g., friendships). Thus, the aforementioned negative other-directed affect typically exhibited by perpetrators of bias which results in social backlash (Czopp, 2019) could be reduced for those sharing a social bond and emotional capital.

Perpetrators who care about getting along with their confronter might exhibit negative self-directed emotional reactions, protecting their confronter from backlash. In nonexistent relationships where there is no goal of restoration or maintenance, the default reaction to the challenge posed by a confronter may rely heavily on negative other-directed emotions and eventually lead to backlash. A component of negative other-directed affect is anger projected toward the confronter. Relationship pairs often share a sense of dependence within their bond influencing them to exhibit hurt feelings rather than anger in the face of a threat (Lemay et al., 2012). Hurt feelings, or the social pain one experiences after relational rejection or devaluation, can be likened with physical pain (Chen et al., 2008), and is most similar to emotions included in negative self-directed affect. Anger, on the other hand, is included in negative other-directed affect. In contrast to hurt feelings, when evoked by relational devaluation, anger is marked by its interpersonally destructive nature (Lemay et al., 2012). Hurt feelings trigger the perpetrator’s
goal to restore social acceptance in constructive ways while anger induces coercion and control, eventually destroying the bond (Murray et al., 2006).

**The Unique Influence of Trust**

As an essential element of human interaction (Rotter, 1971), trust might influence the outcome of interpersonal confrontations of anti-Black racism. Schweers Cook (2005) describes trust as an essential social lubricant, a concept that contributes to the social order. Trust has been defined in a variety of ways, often depending upon the field of study (Rousseau et al., 1998). The ubiquitous characteristics of trust include the expectation of benevolence in another’s motives and intentions, as well as a prediction that others will act in one’s best interest (Lewicki et al., 1998; Rotter, 1971; Deutsch, 1958).

The origin of trust has been a topic of debate. Some hypothesize that trust develops incrementally over time with increased exposure to a specific risky situation with another person (Kramer, 1999). It follows that individuals in established relationships that have withstood relational threats in the past would maintain trust in their partners after experiencing a confrontation for racism. Other research posits that trust develops precipitously and swiftly without the promise of future interactions (Huang & Murnighan, 2010) as well as unexpectedly and rapidly upon first impression (McKnight et al., 1998; Weber et al., 2004). These researchers suggest that an existing personal history or relationship with an interaction partner is not necessary for someone to view the other as trustworthy. Importantly, the research agrees that in order for individuals to evaluate their trust in another, the situation must be inherently risky whereby the intentions of the other person are unknown (Riegelsberger et al., 2003).

There are a number of ways to measure trust, some tied to individual differences (Erikson, 1964; Couch & Jones, 1997), and others to relationship history (Rempel et al., 1985;
Simpson, 2007; Robbins, 2016). Most of the literature studying the outcomes of trust have used relational trust scales due to their emphasis on investigating trust between people (e.g., intimate partners) who share prior trust-relevant experiences. Because trust can spontaneously arise within and outside of relationships, it is important to look at trust from an interpersonal dynamic’s perspective (Goto, 1996). That way, trust can apply to complete strangers or individuals who share a tenuous social bond. The definition of trust encompasses the expectation and prediction of another’s benevolent actions and intentions. Montoya and Pittinksy (2011) have measured the expectation of benevolence between unknown groups with no prior relationship history using the outgroup trust scale. The authors found liking and cooperation between groups as an outcome of increased trust. Similarly, Robinson (1996) used a workplace measure of trust to determine the impact of a psychological contract breach on subsequent employee cooperation with a company. The measure taps into the prediction of another’s benevolence and integrity, without specifically addressing employees’ relationship history. It is important to consider measures like these that remove history when exploring the interpersonal dynamics of trust because the concept is not solely associated with relationship pairs.

Regardless of how it develops, trust influences the results of conflict, making it an understudied variable in the confrontation literature. The present research tested the extent to which trust shaped the outcomes of confrontations for racism in terms of attitude and behavioral changes. Related research shows that greater feelings of trust for government authorities has been linked to increased cooperation and attitudinal alignment with policy proposals (Ross et al., 2014). Likewise, trust may prove to be essential in aligning a perpetrator’s attitude with their confronter’s and engendering cooperation with a request for change.
In intimate relationships, increased trust for a partner who identifies a goal that requires the other partner’s sacrifice is associated with greater accommodation during a discussion of that goal (Shallcross & Simpson, 2012). A perpetrator who trusts their socially bonded confronter might be more open to a conversation about compliance with egalitarian norms moving forward. Importantly, a meta-analysis revealed that the association between trust and behavioral cooperation is strong no matter the length of the relationship history of the individuals involved in a situation (Balliet & Van Lange, 2013). Thus, a perpetrator’s attitudes and behavior subsequent to a confrontation might depend on their level of trust for the confronter regardless of how long they have been acquainted (see Hypothesis 4).

There is evidence to suggest that trust may also influence the emotional outcomes of interactions that involve conflict. Campbell and colleagues (2010) suggest that lower levels of trust are related to more negative evaluations of the overall experience of relationship-based conflict, as well as increased negative externalizing behaviors. In intimate relationships, for example, partners with low trust generate decreased feelings of closeness and forgiveness following conflict, and an increased likelihood of feeling contempt for their partner (Kim et al., 2015). Individuals low in trust tend to attempt to suppress negative emotions after relational conflicts, which results in an increase of negative mood (Righetti et al., 2015). With that, people low in trust may succumb to increased negative other-directed affect after a confrontation which may cause them to engage in backlash. In this way, trust for the confronter may influence the emotional outcomes of conflict in the form of a confrontation.

Due to the fact that trust can be disconnected from relationship standing (Rousseau et al., 1998), I treated trust as a moderator. Because it is possible that people trust others that they do not know well (Huang & Murnighan, 2010; McKnight et al., 1998; Weber et al., 2004), I
hypothesized that trust would moderate the association of one’s motivation to get along with a confronter on negative self- and other-directed affect (see Hypothesis 1b). The outcomes of confrontation (i.e., backlash and prejudice reduction) may also hinge on trust (Brewer, 1981). For people wishing to maintain a relationship, increased feelings of trust might produce more internalized, negative self-directed emotions after a racial confrontation, while trust may have no influence on negative other-directed emotions. In keeping with the aforementioned intergroup sensitivity effect and social identity theory (Tajfel, 1978; Tajfel & Turner, 1979), White people who experience a confrontation of racism by a White stranger might trust that their confronter is acting in their best interest. Because of their shared racial group membership, White people might view their fellow ingroup member as attempting to help them adhere to group norms of egalitarianism, thereby trusting them regardless of a pre-existing relationship. However, the existence of a relational sensitivity effect might dictate that White perpetrators confronted for racism by White strangers with whom no social bond exists are less likely to internalize the message and feel negative self-directed affect, even with increased trust. In this way, people low in a desire to maintain a relationship might not experience any influence of trust on negative self-directed emotions, but increased trust could decrease negative other-directed affect. Just as trust may moderate the impact of one’s motivation to get along with a confronter on negative self- and other-directed emotions, so too could it moderate the impact of one’s motivation to get along with a confronter on prejudice reduction and backlash, respectively (see Hypothesis 4) in an identical way.

In summary, the rose-colored lenses through which people see close others might contribute to positive illusions of socially bonded others, even those with a tenuous relationship.
under which these tough conversations about race could begin at very little social cost. The combination of social bonds and trust might equate to a perfect storm, allowing White confronters to escape backlash and challenge their White close others’ racial bias to reduce prejudice.
CHAPTER FOUR

STUDY DESIGN

Confronting anti-Black racism has the potential to reduce its occurrence (Czopp et al., 2006), especially when it is performed by a White ally (Czopp & Monteith, 2003; Rasinski & Czopp, 2010; Gulker et al., 2013; Schultz & Maddox, 2013). However, many White people report fearing the potential for social backlash associated with confronting racism and therefore opt out of confronting as a response to bias (Eliezer & Major, 2012; Kaiser & Major, 2006). The literature investigating the outcomes of confronting bias has focused primarily on stranger interactions, often ignoring the outcomes of confrontations that occur between people who are motivated to get along with their confronter. This research tests whether the outcomes (i.e., perpetrators’ prejudice reduction and social backlash toward the confronter) differ for confrontations of anti-Black racism between White people who are motivated to get along versus those who are low in that motivation. Additionally, I tested whether trust moderated the strength of the observed associations as research to date has not yet examined the role of trust in confrontations of anti-Black racism.

Hypotheses

**Hypothesis 1:** (a) This exploratory hypothesis will determine whether trust measured either before or after the racial (v. rude) confrontation will moderate the strength of the association between affiliative motivation and negative self- and other-directed emotions (Study 1).
Following a racial confrontation, trust for the confronter will moderate the strength of the association between affiliative motivation and negative self- and other-directed emotions (see also Figure 1, Path A; Study 1 and 2). When affiliative motivation is high, as trust for the confronter increases, negative *self-directed* affect will increase, but the impact of trust on negative *other-directed* affect will not be significant. When affiliative motivation is low, the impact of trust on negative *self-directed* affect will not be significant, but as trust for the confronter increases, negative *other-directed* affect will decrease.

Following a rude confrontation, trust for the confronter will not significantly impact the association between affiliative motivation and negative emotionality.

**Hypothesis 2:** Replicating past research, negative emotionality will mediate the relationship between confrontation content and the outcomes of prejudice reduction and backlash (Study 2). Racial (v. rude) confrontations will increase both negative self- and other-directed emotions. As negative self-directed emotions increase, prejudice and backlash will decrease. As negative other-directed emotions increase, prejudice and backlash will increase.

**Hypothesis 3:** Extending past research, I tested whether negative emotionality mediated the association between affiliative motivation with the confronter and prejudice reduction and backlash following a racial confrontation (see also Figure 1, Paths B & C; Study 2).

After a racial confrontation, participants with high affiliative motivation will experience increased negative self-directed affect compared to those with low affiliative motivation, which will increase prejudice reduction and decrease backlash. Low affiliative motivation will not significantly impact negative self-directed affect, which will not significantly impact prejudice reduction or backlash.
After a racial confrontation, participants with low affiliative motivation will experience increased negative other-directed affect than those with high affiliative motivation, which will decrease prejudice reduction and increase backlash. High affiliative motivation will not significantly impact negative other-directed affect, which will not significantly impact prejudice reduction or backlash.

Following a rude confrontation, affiliative motivation will not significantly impact negative emotionality, resulting in a null effect on prejudice reduction and backlash.

**Hypothesis 4**: Trust will moderate the association between affiliative motivation with the confronter and the outcomes of prejudice reduction and backlash (see also Figure 1, Path D; Study 2). Following a racial confrontation, as trust increases, so will the magnitude of the effect of affiliative motivation on prejudice reduction and backlash. Following a rude confrontation, trust will not significantly impact the strength of the association between affiliative motivation on prejudice reduction and backlash.
Figure 1. Moderated mediation of the indirect effects of NegSelf and NegOther on the association between affiliative motivation and prejudice reduction and backlash by trust for racial (v. rude) confrontations. Path letters correspond to hypotheses 1, 3, and 4.

**Study 1**

The purpose of Study 1 was to test hypotheses 1a and 1b, that trust for the confronter measured either before or after a racial (v. rude) confrontation moderated the strength of the association between affiliative motivation with the confronter and negative self- and other-directed emotions. Because trust has not yet been examined in the literature on confrontation,
this study established whether trust measured before or after a situation involving a confrontation influenced the association between affiliative motivation with the confronter and their emotional reactions (hypothesis 1a). Trust measured before a confrontation could influence the way a perpetrator of racism responds to a confrontation. Research has found that priming trust goals before delivering a message increased participants’ positive evaluation of the message and behavioral intentions to comply (Légal et al., 2012). In this way, trust measured before a confrontation might change the way participants interpret and internalize the message than if trust-relevant words were not presented beforehand. Along the same lines, trust measured after a confrontation could influence participants’ assessment of their emotional reactions. Immediately inquiring about participants’ trust in the messenger after receiving feedback could amplify or reduce negative feelings, diminishing the accuracy of the subsequent negative emotionality measure.

Moreover, research has determined that a desire to get along with or maintain a relationship with another elicits more restorative negative feelings after conflict (i.e., negative self-directed affect) and less destructive feelings (i.e., negative other-direct affect; Feeney & Lemay, 2012; Lemay et al., 2012). Additionally, trust and negative other-directed emotions are often negatively correlated (Kim et al., 2015; Righetti et al., 2015). Taken together, this study was the first to account for the roles of both affiliative motivation and trust in the outcomes of racial confrontation. I tested the hypothesis that when affiliative motivation was high, as trust for the confronter increased, negative self-directed affect increased and the impact of trust on negative other-directed affect was not significant. On the other hand, when affiliative motivation was low, as trust for the confronter increased, negative other-directed affect decreased and the impact of trust on negative self-directed affect was not significant (hypothesis 1b).
Study 1 Method

Design

A 2(affiliative motivation: low v. high) x 2(confrontation: racial v. rude) x 2(trust: before confrontation v. after confrontation) between-subjects design measured emotional reactions to a confrontation of anti-Black bias.

Prospective Power Analysis

Based on a small effect size ($f = .15$) found in prior research (e.g., Czopp et al., 2006) and to account for the expected interactions, results of a prospective power analysis conducted using G*Power 3.1 (Erdfelder et al., 1996; Faul et al., 2007) indicated that 350 participants would be an adequate sample size to achieve 80% power to detect effects. In anticipation that some participants would not meet the study’s inclusion criteria and/or complete the study in good faith, additional participants were recruited.

Participants

A total of 538 adult, U.S. citizens (351 women, 181 men, and 6 non-binary) between 18 and 83 years old ($M = 45.75$, $SD = 14.31$) who were recruited through Amazon’s Mechanical Turk crowdsourcing website (see Appendix A) passed the pre-determined screening checks (see Data Screening below). All participants were compensated $0.50 for the completion of the tasks, which took less than 30 minutes, and a bonus incentive of $0.50 for passing 50% of the attention checks with no missing data.

Procedure

All materials were presented to the participants through Qualtrics survey software. Participants indicated their consent by clicking a link to proceed from the informed consent document (see Appendix B) to subsequent pages of the survey. After reading the cover story,
participants selected an avatar and were randomly assigned to either experience the low or high affiliative motivation manipulation. After the affiliative motivation manipulation, half of the participants were randomly assigned to complete the trust measure while the other half completed the measure after the confrontation. All participants completed the photo-inference task and received feedback from their supposed interaction partner about their performance on the task. Participants were randomly assigned to receive a racial or rude confrontation for their performance on the task. Finally, participants finished the study by filling out the emotional reaction questionnaire, demographic items, and a hypothesis guessing, engagement and debriefing exercise.

**Materials and Measures**

The cover story for this study led participants to believe they were participating in research focused on determining their reasoning ability (see Appendix C). As is typical for research conducted online (Abbey & Meloy, 2017), attention checks asking participants to select particular responses or complete simple math equations were scattered throughout the procedure to ensure participants completed the study in good faith.

**Avatar selection.** Participants were asked to select from four nearly identical avatars, all of which exhibited a light skin tone (see Appendix C). Their selection appeared each time they engaged in the chat with their supposed interaction partner. The other participant’s avatar was not available as a selection to the participant to ensure that similarity was not primed, but both avatars had light skin tones to signal to the participant that their interaction partner was White.

**Affiliative motivation manipulation.** Participants were randomly assigned to either experience low or high affiliative motivation with their supposed interaction partner following a similar procedure to that used by Sinclair and colleagues (2005; see Appendix D). After reading
the cover story, participants were told that they would be interacting with another participant simultaneously completing the reasoning ability study. All participants were shown a screen purportedly connecting them with their interaction partner. Once the connection occurred, participants were instructed to introduce themselves to their partner by indicating their name, hometown, birthday, and favorite color. In the high affiliative motivation condition, participants were told in the response by their interaction partner that they shared the same birthday. The following survey prompt indicated that they would interact with their partner for the remainder of the study (i.e., 30 minutes). In the low affiliative motivation condition, participants were not told that they shared a birthday with their partner during the introductions and the subsequent survey prompt indicated that they would interact for only five minutes over the remainder of the study.

In order to ascertain whether the affiliative motivation manipulation was salient, participants were asked, “What, if anything, did you share in common with your partner?” (i.e., hometown, birthday, favorite food, none of the above). Additionally, participants were asked, “How long will you interact with your partner for this study” (i.e., 30 minutes, 5 minutes, I don’t remember). To test whether the affiliative motivation manipulation was effective, participants filled out a short questionnaire indicating how similar and close they felt to the other participant on a scale from 1 (not at all) to 7 (very much). The two target items were mixed in with questions such as, “How alert do you feel right now?” and “How many online studies have you participated in?” as to disguise the purpose of the measure (see Appendix E).

Trust. Participants were randomly assigned to receive the trust measure either before or after the confrontation using a modified outgroup trust scale (Montoya & Pittinsky, 2011), which contained 8 items (α = .89; e.g., “I believe that the other participant will look out for my
interests”) answered from 1 (not at all) to 7 (very much). Additionally, participants completed a modified version of Robinson’s (1996) workplace trust scale, which contained 7 items (α = .86; e.g., “I can expect the other participant to treat me in a consistent and predictable fashion”) answered from 1 (strongly disagree) to 7 (strongly agree; see Appendix F).

**Photo-inference task.** Participants were told that their interaction partner would be able to see their responses as they completed a photo-description inference task (Czopp et al., 2006). This task elicited stereotypic responding. On each trial of the task (see Appendix G), participants saw a photograph of a person along with a short description and they were instructed to type a 1- or 2-word inference about that person’s job or hobby. For example, a picture of a White man with the description, “This person can be found in a theater,” might elicit the response, “Actor”. The critical trials contained photographs that typically elicit stereotypic responses. For example, a photograph of a Black man along with a description (e.g., “This person can be found behind bars”) most often evokes a stereotypic response (e.g., “Criminal”) compared to a non-stereotypic response (e.g., “Bartender”).

**Confrontation.** The interaction partner who was purportedly watching the task commented on the participant’s performance. The feedback provided by the alleged partner was similar to that used in Czopp et al. (2006). Participants were randomly assigned to either receive racial or rude content in the confrontation disguised as feedback (see Appendix H). The racial confrontation read, “You should really try to think about Black people in other ways that are less prejudiced. It just seems that you are some kind of racist. You know what I mean?”, and the rude confrontation read, “Were you even paying attention to what you were doing? It was impossible for me to follow you. You know what I mean?” After receiving the confrontation, participants
had a chance to respond, and the survey screen advanced them along to the next prompt once their response was recorded.

**Emotional reactions.** Finally, participants completed the NegSelf/NegOther questionnaire (Devine et al., 1991; see Appendix I). The scale ranged from 1 (does not apply at all) to 7 (applies very much) and included 11 items. Participants indicated the extent to which they were feeling 8 negative self-directed emotions (α = .96; e.g., guilty, regretful), and 3 negative other-directed emotions (α = .89; e.g., angry at others, irritated with others).

**Demographics.** All participants responded to a series of demographic items asking for their guess of the other participant’s gender and race. They also indicated their own age, gender, and race (see Appendix J).

**Hypothesis guess and performance evaluation.** Finally, participants submitted their guess for the hypothesis of this study and indicated how much they believed that another person was truly connected with them from 1 (not at all) to 7 (very much). Participants were also asked to evaluate their own performance during the study with 2 items (r = 0.47, p < .001), “I took this study seriously” and “You should use my data in your analyses”, from 1 (strongly disagree) to 7 (strongly agree; see Appendix K).

**Debriefing.** Finally, all participants were presented with a debriefing document (see Appendix L) before inputting their Worker ID for compensation.

**Study 1 Results**

**Data Screening**

The total sample was comprised of 747 participants recruited between January 5, 2021, and February 2, 2021. The data were screened in a series of 3 steps. Of the total sample, 575 participants were White identified (step 1) and passed 50% or more of the 5 attention checks.
scattered throughout the task (step 2). Of these participants, 538 completed the procedure in good faith by indicating their birthday in the introduction – an essential component of the affiliative motivation manipulation (step 3; see Figure 2, and Appendix M for more details).

Figure 2. Details of the data screening process for Study 1.

Manipulation Checks

With these exclusions, I reviewed the zero-order correlation of participants’ feelings of closeness and similarity with their interaction partner. The two variables were highly correlated, $r = 0.78$, $p < .001$ (see Table 1), so I combined them to create a variable representing the
connection participants felt with their interaction partner. An independent samples t-test confirmed, \( t(536) = -2.57, p < .05; d = 0.23 \), that participants in the high affiliative motivation condition (\( M = 3.71, SD = 1.48 \)) felt more connected to their interaction partner compared to those in the low affiliative motivation condition (\( M = 3.38, SD = 1.45 \)). There was no difference in alertness, \( t(536) = 0.57, p = .57; d = 0.05 \), or the number of studies participants in either affiliative motivation condition previously completed, \( t(536) = -0.55, p = .58; d = 0.05 \).

**Descriptive Statistics**

After reviewing the bivariate correlations for all the variables of interest (see Table 1), I combined the two trust measures (i.e., outgroup trust and workplace trust) as they were highly correlated, \( r = 0.86, p < .001 \). Affiliative motivation and the overall trust measure were not correlated, \( r = 0.002 p = .97 \), and there was no difference between participants in the high (\( M = 3.99, SD = 1.30 \)) or low (\( M = 3.98, SD = 1.28 \)) affiliative motivation condition in terms of trusting their interaction partner, \( t(536) = -0.04, p = .97; d = 0.004 \).
Table 1. Bivariate correlation matrix detailing the associations between variables of interest.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affiliative Motivation</td>
<td>--</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>2. Confrontation Type</td>
<td>--</td>
<td>-.02</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>3. Trust Condition</td>
<td>--</td>
<td>.02</td>
<td>-.004</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>4. Combined Trust Scale</td>
<td>3.99</td>
<td>.002</td>
<td>-.06</td>
<td>.58**</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Outgroup Trust</td>
<td>3.84</td>
<td>.003</td>
<td>-.05</td>
<td>-.58**</td>
<td>.96**</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Workplace Trust</td>
<td>4.13</td>
<td>.001</td>
<td>-.06</td>
<td>-.53**</td>
<td>.97**</td>
<td>.86**</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. NegSelf</td>
<td>1.71</td>
<td>.03</td>
<td>-.002</td>
<td>.04</td>
<td>.06</td>
<td>.06</td>
<td>.05</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NegOther</td>
<td>2.48</td>
<td>-.01</td>
<td>.05</td>
<td>-.08</td>
<td>-.04</td>
<td>-.04</td>
<td>.36**</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Closeness</td>
<td>3.48</td>
<td>.08**</td>
<td>.04</td>
<td>-.004</td>
<td>.24**</td>
<td>.22**</td>
<td>.24**</td>
<td>.17**</td>
<td>.09</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Similarity</td>
<td>3.67</td>
<td>.13**</td>
<td>.04</td>
<td>.04</td>
<td>.20**</td>
<td>.20**</td>
<td>.20**</td>
<td>.16**</td>
<td>.08</td>
<td>.78**</td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* *p < .05, **p < .01, *p = .05

**Hypothesis 1A**

First, I tested hypothesis 1a to determine whether trust measured before or after the racial (v. rude) confrontation moderated the strength of the association between affiliative motivation and negative self- and other-directed emotions. I started by conducting two three-way ANOVAs with affiliative motivation (high v. low), confrontation type (racial v. rude), and trust time point (before v. after confrontation) as fixed factors, and NegSelf and NegOther, respectively, as the dependent variables. Any significant two- or three-way interactions were examined to ascertain the trust time point driving the effect. Additionally, all analyses of variance controlled for participants’ evaluation of how seriously they took the study.
With NegSelf entered as the dependent variable, the ANOVA indicated no main effects of affiliative motivation, confrontation type, or trust time point. There were no significant two-or three-way interactions between affiliative motivation and trust time point, confrontation type and trust time point, or affiliative motivation, confrontation type and trust time point (see Table 2).

With NegOther entered as the dependent variable, results revealed no main effects of affiliative motivation or confrontation type, but a marginally significant main effect of trust time point (see Table 2). Participants who completed the trust measures before receiving a confrontation reported more negative other-directed emotions ($M = 2.62$, $SD = 1.84$) than those who completed the trust measures after the confrontation ($M = 2.34$, $SD = 1.57$). While there were no two-way interactions of affiliative motivation and confrontation type, or affiliative motivation and trust time point, there was a significant interaction of confrontation type and trust time point.

To probe this interaction, I tested the simple main effects. When trust was measured before the confrontation, $F(1, 529) = 0.48, p = .49, \eta^2_p = .001$, there was no difference in negative other-directed emotions between those who received a racial versus rude confrontation. However, when trust was measured after the confrontation, $F(1, 529) = 6.77, p < .05, \eta^2_p = .01$, participants who received a racial confrontation felt more negative other-directed emotions ($M = 2.59$, $SD = 1.72$) than those who received a rude confrontation ($M = 2.07$, $SD = 1.36$). The three-way interaction of affiliative motivation, confrontation type, and trust time point was not significant (see Table 2). This provided support for measuring trust after the confrontation.
Table 2. ANOVA results detailing the effects of affiliative motivation, confrontation type, and trust on negative self- and other-directed emotions.

<table>
<thead>
<tr>
<th>DV: NegSelf</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.45</td>
<td>0.32</td>
<td>.57</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>0.11</td>
<td>0.08</td>
<td>.78</td>
<td>.001</td>
</tr>
<tr>
<td>Trust (C)</td>
<td>1</td>
<td>1.95</td>
<td>1.37</td>
<td>.24</td>
<td>.003</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.32</td>
<td>0.22</td>
<td>.64</td>
<td>.001</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>1.42</td>
<td>1.00</td>
<td>.32</td>
<td>.002</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>0.28</td>
<td>0.20</td>
<td>.66</td>
<td>.001</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>0.06</td>
<td>0.04</td>
<td>.84</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>529</td>
<td>1.42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: NegOther</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.17</td>
<td>0.06</td>
<td>.81</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>5.64</td>
<td>1.93</td>
<td>.17</td>
<td>.004</td>
</tr>
<tr>
<td>Trust (C)</td>
<td>1</td>
<td>10.95</td>
<td>3.74</td>
<td>.05*</td>
<td>.007</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.33</td>
<td>0.11</td>
<td>.74</td>
<td>.001</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>0.39</td>
<td>0.13</td>
<td>.72</td>
<td>.001</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>16.23</td>
<td>5.54</td>
<td>.02*</td>
<td>.010</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>1.21</td>
<td>0.41</td>
<td>.52</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>529</td>
<td>2.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05, +p = .05

Hypothesis 1B

I next tested hypothesis 1b (i.e., when affiliative motivation was high, as trust for the confronter increased, negative self-directed affect increased, but there was no impact on negative other-directed affect; when affiliative motivation was low, as trust for the confronter increased, negative other-directed affect decreased, but there was no impact of trust on negative self-directed affect).

I used the PROCESS macro for SPSS which estimates a process model for moderation (Model 2; Preacher et al., 2007) with bootstrapping (n = 5000). Before running the analyses, I separated cases by trust time point (before v. after the confrontation) to examine the results within each condition. In this analysis, affiliative motivation (low = 0 v. high = 1) served as the
independent variable, confrontation type (racial = 1 v. rude = 0) and the continuous measure of trust were moderators, and negative self- and other-directed emotions served as dependent variables.

When trust was measured after the confrontation, the continuous trust variable did not moderate the association between affiliative motivation and negative self-directed emotions, $F(1, 256) = 0.36, p = .55, R^2 = .001$, in either the racial or rude confrontation condition, $F(1, 256) = 0.12, p = .73, R^2 = .001$. However, the results showed a marginally significant effect such that trust measured after the confrontation moderated the association between affiliative motivation and negative other-directed emotions, $F(1, 256) = 3.66, p = .06, R^2 = .01$, regardless of confrontation type, $F(1, 256) = 0.27, p = .60, R^2 = .01$. The results suggest that regardless of whether participants are confronted for racism or rudely, when affiliative motivation is high, the impact of trust on negative other-directed affect is not significant but when affiliative motivation is low, as trust for the confirter increases, negative other-directed affect decreases. These results partially support hypothesis 1b (see Table 3 and Figure 3).

When trust was measured before the confrontation, the continuous trust variable did not moderate the association between affiliative motivation and negative self-directed emotions, $F(1, 270) = 0.86, p = .36, R^2 = .003$, in either the racial or rude confrontation condition, $F(1, 270) = 0.02, p = .88, R^2 = .001$. Likewise, trust did not moderate the association between affiliative motivation and negative other-directed emotions, $F(1, 270) = 0.00, p = .98, R^2 = .001$, in either confrontation condition, $F(1, 270) = 0.02, p = .90, R^2 = .001$.

1. See Appendix N for more information about the results of these analyses when I adopted a more stringent data screening process.
Table 3. PROCESS Model 2 results for Hypothesis 1a depicting the interaction between affiliative motivation and trust measured before and after a confrontation on negative self- and other-directed emotions.

<table>
<thead>
<tr>
<th>Before the Confrontation</th>
<th>Negative Self-Directed Emotions</th>
<th>Negative Other-Directed Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE) [LLCI, ULCI]</td>
<td>b (SE) [LLCI, ULCI]</td>
</tr>
<tr>
<td>Affiliative Motivation (AM)</td>
<td>-0.57 (0.83) [-2.21, 1.08]</td>
<td>-0.09 (1.27) [-2.59, 2.41]</td>
</tr>
<tr>
<td>Confrontation</td>
<td>-0.02 (0.23) [-0.47, 0.44]</td>
<td>-0.17 (0.35) [-0.86, 0.53]</td>
</tr>
<tr>
<td>AM x Confrontation</td>
<td>-0.04 (0.30) [-0.64, 0.55]</td>
<td>0.06 (0.46) [-0.84, 0.96]</td>
</tr>
<tr>
<td>Trust (continuous)</td>
<td>-0.08 (0.13) [-0.34, 0.18]</td>
<td>0.07 (0.20) [-0.32, 0.47]</td>
</tr>
<tr>
<td>AM x Trust</td>
<td>0.16 (0.17) [-0.18 – 0.49]</td>
<td>-0.01 (0.26) [-0.51, 0.50]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>After the Confrontation</th>
<th>Negative Self-Directed Emotions</th>
<th>Negative Other-Directed Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE) [LLCI, ULCI]</td>
<td>b (SE) [LLCI, ULCI]</td>
</tr>
<tr>
<td>Affiliative Motivation (AM)</td>
<td>-0.21 (0.46) [-1.12, 0.71]</td>
<td>-0.83 (0.59) [-2.00, 0.34]</td>
</tr>
<tr>
<td>Confrontation</td>
<td>0.13 (0.23) [-0.33, 0.58]</td>
<td>0.57 (0.30) [-0.01, 1.15]</td>
</tr>
<tr>
<td>AM x Confrontation</td>
<td>-0.10 (0.30) [-0.69, 0.48]</td>
<td>-0.20 (0.38) [-0.95, 0.55]</td>
</tr>
<tr>
<td>Trust (continuous)</td>
<td>0.12 (0.10) [-0.07, 0.32]</td>
<td>-0.50 (0.13) [-0.75, -0.25]</td>
</tr>
<tr>
<td>AM x Trust</td>
<td>0.08 (0.13) [-0.17 – 0.32]</td>
<td>0.31 (0.16) [-0.01, 0.62]</td>
</tr>
</tbody>
</table>

**Conditional effects of AM x Trust measured after the confrontation on NegOther at three levels of trust collapsed across confrontation condition.** Analyses use continuous variables; effects below are one standard deviation +/- the mean.

<table>
<thead>
<tr>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low ($M = 1.75$)</td>
<td>-0.41 (0.30) [-1.00, 0.19]</td>
</tr>
<tr>
<td></td>
<td>Medium ($M = 3.32$)</td>
<td>0.09 (0.19) [-0.29, 0.47]</td>
</tr>
<tr>
<td></td>
<td>High ($M = 4.34$)</td>
<td>0.42 (0.26) [-0.10, 0.93]</td>
</tr>
</tbody>
</table>

*Note: $^a p = 0.56$, $^b p = 0.54$, $^** p < .01$*
Study 1 Discussion

The results of Study 1 support the general notion that trust influences negative emotionality following a confrontation for anti-Black racism, especially for perpetrators whose desire to get along with their confronter is low. Trust influenced the association between affiliative motivation and negative other-directed emotions (hypothesis 1b). As predicted, as trust for their confronter increased people who were low in the desire to get along with their confronter reported less negative other-directed emotions, like anger or annoyance. Past research has found that negative other-directed affect, like anger, can be destructive to interpersonal relationships (Lemay et al., 2012; Murray et al., 2006), suggesting that those with a desire to get along with another might be less likely to exhibit negative-other directed emotions. However, this is the first research to examine the impact of trust on one’s social bond and negative
emotionality following conflict. In this study, trust influenced reports of negative other-directed affect for those individuals who were low in the desire to get along with their confronter, reducing their negative other-directed emotions to levels reported by those with a high desire to get along with their confronter. Following from prior research suggesting that trust is negatively associated with negative emotions, like contempt (Kim et al., 2015), individuals with low affiliative motivation who felt increased trust were less likely to report negative-other directed emotions following a confrontation compared to those who reported lower levels of trust. Contrary to hypotheses, the effect remained marginally significant (i.e., falling below conventional standards of significance but still providing reasonably interpretable results) for participants in both the racial and rude confrontation conditions, perhaps due to both confrontations equally signaling interpersonal conflict.

In this study trust was not merely a proxy for high affiliative motivation as the two variables were not significantly correlated. Instead, trust mitigated negative other-directed affect for participants with low affiliative motivation, protecting the confronter from the impact of the perpetrator’s negative emotions. Participants in the high affiliative motivation condition did not exhibit a change in negative other-directed emotions as trust increased; their overall levels of negative other-directed emotions remained lower than participants in the low affiliative motivation condition. These results are promising as prior research has shown that decreased negative other-directed emotions are often linked to less backlash for confronters (Monteith et al., 1993; Czopp et al., 2006). Study 2 will extend the current research to include a measure of backlash to determine if the effect translates to a decrease in backlash.

Contrary to predictions, trust did not moderate the association between affiliative motivation and negative self-directed emotions following a confrontation (hypothesis 1b).
Negative self-directed emotions, like guilt, are often reparative to social bonds and can be adopted by individuals wishing to maintain a connection (Baumeister et al., 1994). Because this study was conducted in an online setting, the predicted effect of trust on the association between affiliative motivation and negative self-directed emotions may have been diffused by the improbability of a continuing bond with the confronter. The affiliative motivation manipulation might have been enough to induce avoidance of further damaging a social bond with interpersonally destructive emotions (i.e., negative other-directed affect), but it may not have been strong enough to elicit bond reparative motivations in this context (i.e., negative self-directed affect).

These results only held when trust was measured after (v. before) the perpetrator received the confrontation (exploratory hypothesis 1a). Interestingly, participants who completed the trust measures before the confrontation reported significantly more negative other-directed emotions than those who completed the measures afterward, and this was equally true in both the racial and rude confrontation conditions. Overall levels of negative other-directed emotions were lower when trust was measured after the confrontation, perhaps because measuring trust after the confrontation dampened the emotional responses of participants due to the time between the confrontation and completing the negative emotionality questionnaires. However, trust measured after the confrontation increased negative-other directed emotions for participants confronted for racism versus those confronted rudely, which follows from previous research finding that racial confrontations elicit more negative other-directed emotions (Czopp et al., 2006). Additionally, trust is typically only required in the presence of risk (Riegelsberger et al., 2003). In other words, evaluating trust after a confrontation is more likely to occur in real life situations, as there is risk in not knowing the confronter’s intentions. Thus, measuring trust before a confrontation not only
produced null results in terms of the hypotheses, but it is unlikely that individuals evaluate their trust in another without experiencing the risk of having to infer their intentions after conflict.

Although the affiliative motivation manipulation was effective in that those in the high (v. low) affiliative motivation condition felt more connected to their interaction partner, there was a great deal of variability in this online study. The analyses using the more comprehensive sample capitalized on the power of random assignment, provided more power to detect effects, and provided an answer for hypothesis 1a (i.e., to measure trust after the confrontation). Unfortunately, excluding those participants who did not pass the manipulation check indicating their feelings of closeness with their interaction partner – an interpretable report as to the effectiveness of the manipulation – would decrease the sample by over half. In the end, convincing participants that they were indeed interacting online with another person proved difficult, likely limiting the impact of the affiliative motivation manipulation. Because it was difficult to manipulate affiliative motivation and achieve the intended effect online, Study 2 used a more detailed introduction and a script that mirrored prior online manipulations of affiliative motivation (Huntsinger et al., 2016).

**Study 2**

The purpose of this study was to replicate the effect of trust on the association between affiliative motivation and negative emotionality following a confrontation (hypothesis 1b). It further extended Study 1’s novel findings to determine the effect of trust on backlash and prejudice reduction outcomes (hypothesis 4). Drawing from the logic of hypothesis 1, I predicted that after a racial confrontation, when affiliative motivation was high, as trust for the confronter increased, prejudice reduction increased and the impact of trust on backlash was not significant.
When affiliative motivation was low, as trust for the confronter increased, backlash decreased and the impact of trust on prejudice reduction was not significant.

Additionally, Study 2 sought to replicate prior studies (e.g., Czopp et al., 2006; Monteith et al., 1993) and tested hypothesis 2 that negative emotionality mediates the association between the content of a confrontation (racial or rude) and the backlash and prejudice reduction outcomes. I also tested hypothesis 3, that negative emotionality mediates the association between one’s relationship with the confronter and prejudice reduction and backlash outcomes for a racial, but not rude, confrontation.

Past research demonstrated that guilt, or negative self-directed affect, was more common and stronger after a threat to close relationships compared to distant relationships (Baumeister et al., 1994). Thus, I predicted that after a racial confrontation, participants with high affiliative motivation would experience increased negative self-directed affect compared to those with low affiliative motivation, which would increase prejudice reduction and decrease backlash. In contrast, individuals with low (v. high) affiliative motivation tend to externalize and act aggressively (Hill & Werner, 2006; Delveaux & Daniels, 2000) after a threat. Thus, I predicted that after a racial confrontation, participants with low affiliative motivation would experience increased negative other-directed affect than those with high affiliative motivation, which would decrease prejudice reduction and increase backlash.

**Study 2 Method**

**Design**

A 2(affiliative motivation: high v. low) x 2(confrontation: racial v. rude) between-subjects design measured trust, emotional reactions, backlash and prejudice reduction.
Prospective Power Analysis

Based on a small effect size ($f = .15$) found in prior research (e.g., Czopp et al., 2006) and to account for the expected interactions, results of a prospective power analysis conducted using G*Power 3.1 (Erdfelder et al., 1996; Faul et al., 2007) indicated that 500 participants would be an adequate sample size to achieve 80% power to detect effects. In anticipation that some participants would not meet the study’s inclusion criteria and/or complete the study in good faith, additional participants were recruited.

Participants

A total of 869 adult, U.S. citizens (567 women, 295 men, and 7 non-binary) who were between 18 and 89 years old ($M = 41.55, SD = 13.17$) were recruited through Amazon’s Mechanical Turk crowdsourcing website (see Appendix A) passed the predetermined screening checks (see Data Screening below). All participants were compensated $1.00 for the completion of the tasks, which took less than 30 minutes.

Procedure

Participants completed an almost identical procedure to that of study one, such that they provided informed consent (see Appendix O), read the same cover story, and selected an avatar (see Appendix C). The introduction was slightly altered to further persuade participants that they were indeed interacting with another participant (see Appendix P). Participants were asked to choose a number that would be used to determine who would choose topics to be included in the introduction. The topics were presented to the participant, and they were told that the other participant was selected to choose 3 (i.e., hometown, birthday, favorite color). Participants were randomly assigned to either the high or low affiliative motivation manipulation (see Appendix Q), which was slightly amended to include a statement of suspicion to mirror a real interaction
more closely (Huntsinger et al., 2016), and completed the attention and manipulation checks (see Appendix E).

Because the results of the first study indicated that trust measured after the confrontation moderated the association between affiliative motivation and negative other-directed emotions, participants completed the trust measures (see Appendix F) following the photo-inference task (see Appendix G) and after the other participant’s feedback. Participants were randomly assigned to receive either the racial or rude confrontation disguised as feedback from their interaction partner (see Appendix H). They completed the NegSelf/NegOther questionnaire (see Appendix I) followed by the measures described hereafter to assess backlash and prejudice reduction. The study culminated with identical demographics (see Appendix J), hypothesis guessing, and engagement questions (see Appendix K) contained in the first study, as well as a debriefing segment (see Appendix R).

**Materials and Measures**

Materials and measures from study one remained identical in this study (i.e., outgroup trust scale \( \alpha = .83 \), workplace trust scale \( \alpha = .81 \), NegSelf \( \alpha = .94 \), NegOther \( \alpha = .85 \), demographics, hypothesis guess, and performance evaluation \( r = 0.63, p < .001 \)) with the addition of those described below.

**Backlash Against the Confronter.** Participants completed a measure of liking for the confronter as a proxy for backlash that closely resembled the items used by Mallett and Wagner (2011). The scale ranged from 1 (absolutely not) to 5 (absolutely) and included 6 items \( \alpha = .83 \) like, “The other participant is likeable” (see Appendix S).

**Prejudice Reduction.** Participants completed a measure of their attitudes toward Black people (Payne et al., 2010; see Appendix T) by indicating the extent to which they liked or
disliked five different racial groups including Black/African Americans on a scale from 1 (dislike a great deal) to 7 (like a great deal). Then they completed a measure of stereotype endorsement of Black people that asked participants to indicate how well 14 adjectives or phrases ($\alpha = .96$; e.g., dependable, lazy, irresponsible) described Black people on a scale from 1 (not well at all) to 5 (extremely well; Payne et al., 2010; see Appendix U). Finally, intentions to control future bias (Parker et al., 2018; see Appendix V) was measured with 10 items ($\alpha = .96$; e.g., “After what I learned today, I will be more on guard for racially biased behavior) on a scale from 1 (not at all true) to 7 (completely true).

**Study 2 Results**

**Data Screening**

The total sample was comprised of 1,160 participants recruited between February 11, 2021 and March 9, 2021. The data were screened in a series of 3 steps. Of the total sample, 887 participants were White-identified (step 1) and passed 50% or more of the 7 attention checks scattered throughout the task (step 2). Of these participants, 869 completed the procedure in good faith by indicating their birthday in the introduction – an essential component of the affiliative motivation manipulation (step 3; see Figure 4 and Appendix W for more details).
Manipulation Checks

With these exclusions, I reviewed the zero-order correlation of participants’ feelings of closeness and similarity with their interaction partner. As in Study 1, the two variables were highly correlated, $r = 0.77$, $p < .001$ (see Table 4), so I combined them to create a variable representing the connection participants felt with their interaction partner. Contrary to Study 1 results, an independent samples t-test found no difference in participants’ feeling of connection with their interaction partner between the high and low affiliative motivation conditions, $t(867) =$
-0.41, \( p = .68 \); \( d = 0.03 \). Similar to Study 1, participants in the high and low affiliative motivation conditions did not differ in their level of alertness, \( t(867) = 0.63, p = .53 \); \( d = 0.05 \), or number of studies previously completed, \( t(867) = -0.02, p = .99 \); \( d = 0.001 \).

**Descriptive Statistics**

Again, the bivariate correlations for all the variables of interest (see Table 4), indicated that the two trust measures (i.e., outgroup trust and workplace trust) were highly correlated with one another, \( r = 0.79 \), \( p < .001 \), so they were combined in subsequent analyses. As in Study 1, affiliative motivation and the overall trust measure were not correlated, \( r = -0.03 \), \( p = .34 \), and there was no difference between participants in the high or low affiliative motivation condition in terms of trusting their interaction partner, \( t(867) = 0.95, p = .32 \); \( d = .07 \).

The type of confrontation (i.e., racial or rude) was correlated with negative other-directed emotions, \( r = 0.10 \), \( p < .01 \), and an independent samples t-test confirmed, \( t(867) = -2.80, p < .01 \); \( d = .19 \), that participants who experienced a racial confrontation reported more negative other-directed emotions (\( M = 2.65, SD = 1.68 \)) than those confronted rudely (\( M = 2.35, SD = 1.54 \)). Confrontation type was also correlated with backlash, \( r = -0.09 \), \( p < .05 \). Participants confronted for racism liked their confronter less (\( M = 2.23, SD = 0.87 \)) than those confronted rudely (\( M = 2.38, SD = 0.80 \)), \( t(867) = 2.58, p < .05 \); \( d = .18 \). Finally, confrontation type was correlated with stereotype endorsement, \( r = 0.08 \), \( p < .05 \), and intentions to control future bias, \( r = -0.09, p < .01 \). Participants confronted for racism endorsed more positive stereotypes about Black people (\( t(867) = -2.43, p < .05 \); \( d = .17 \); \( M = 3.95, SD = 0.86 \)), but were less likely to report intending to control future bias (\( t(867) = 2.66, p < .01 \); \( d = .18 \); \( M = 2.60, SD = 1.71 \)) compared to those confronted rudely (\( M = 3.80, SD = 0.87 \) and \( M = 2.91, SD = 1.70 \), respectively).
Table 4. Bivariate correlation matrix detailing the associations between variables of interest.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affiliative Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Confront Type</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Closeness</td>
<td>3.12 (.59)</td>
<td>-.01</td>
<td>-.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Similarity</td>
<td>3.45 (1.52)</td>
<td>.04</td>
<td>.02</td>
<td>.77**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Combined Trust Scale</td>
<td>3.08 (1.10)</td>
<td>-.03</td>
<td>-.04</td>
<td>.15**</td>
<td>.18**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Outgroup Trust</td>
<td>2.96 (1.14)</td>
<td>-.04</td>
<td>-.05</td>
<td>.14**</td>
<td>.17**</td>
<td>.94**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Workplace Trust</td>
<td>3.20 (1.20)</td>
<td>-.02</td>
<td>-.03</td>
<td>.14**</td>
<td>.17**</td>
<td>.95**</td>
<td>.79**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NegSelf</td>
<td>1.58 (.99)</td>
<td>-.02</td>
<td>-.01</td>
<td>.11*</td>
<td>.16**</td>
<td>.16**</td>
<td>.15**</td>
<td>.15**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. NegOther</td>
<td>2.49 (1.61)</td>
<td>-.04</td>
<td>.10**</td>
<td>.07*</td>
<td>.07</td>
<td>-.38**</td>
<td>-.37**</td>
<td>-.35**</td>
<td>.28**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Backlash</td>
<td>2.31 (0.84)</td>
<td>-.03</td>
<td>.09*</td>
<td>.19**</td>
<td>.22**</td>
<td>.77**</td>
<td>.74**</td>
<td>.73**</td>
<td>.21**</td>
<td>-.31**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Attitudes Toward Black</td>
<td>5.51 (1.54)</td>
<td>-.04</td>
<td>.06</td>
<td>.11**</td>
<td>.10**</td>
<td>.05</td>
<td>.07</td>
<td>.05</td>
<td>-.07</td>
<td>-.10**</td>
<td>.09**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Stereotype Endorsement</td>
<td>3.87 (.87)</td>
<td>-.01</td>
<td>.08*</td>
<td>.03</td>
<td>.003</td>
<td>.09*</td>
<td>.09**</td>
<td>.07</td>
<td>-.14**</td>
<td>-.12**</td>
<td>.10**</td>
<td>.59**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Intentions to Control</td>
<td>2.76 (1.71)</td>
<td>-.03</td>
<td>.09**</td>
<td>.21**</td>
<td>.18**</td>
<td>.20**</td>
<td>.20**</td>
<td>.18**</td>
<td>.35**</td>
<td>.10**</td>
<td>.24**</td>
<td>.02</td>
<td>-.03</td>
<td></td>
</tr>
</tbody>
</table>

Note: p< .05*, p<.01**

Analyses of Variance

I began exploring patterns in the data that would later inform pieces of the full moderated mediation model (hypotheses 1b-4) by conducting several two-way ANOVAs. The fixed factors for each ANOVA were affiliative motivation (high v. low) and confrontation type (racial v. rude), and the dependent variables were trust, NegSelf, NegOther, backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias, respectively. Any significant two-way interactions were examined to determine whether the results were driven by the racial or rude confrontation condition. Additionally, all analyses controlled for participants’ evaluation of how seriously they took the study (see Table 5).
Trust. The ANOVA indicated no main effects of affiliative motivation or confrontation type on the combined trust measure, nor a significant interaction.

NegSelf. The results showed no significant main effects of affiliative motivation or confrontation type on negative self-directed emotions, nor an interaction.

NegOther. There was no main effect of affiliative motivation, but the effect of confrontation type on negative other-directed emotions was significant. Participants confronted for racism reported more negative other-directed emotions ($M = 2.65, SD = 1.68$) compared to those confronted rudely ($M = 2.35, SD = 1.54$). The interaction of affiliative motivation and confrontation type on negative other-directed emotions was not significant.

Backlash. With backlash entered as the dependent variable, the two-way ANOVA indicated no significant main effect of affiliative motivation, but a significant main effect of confrontation type (see Descriptive Statistics section for more details). The interaction of affiliative motivation and confrontation type on backlash was not significant.

Attitudes Toward Black People. The ANOVA indicated no main effect of affiliative motivation or confrontation type on attitudes toward Black people, nor an interaction.

Stereotype Endorsement. The two-way ANOVA for stereotype endorsement showed no main effect of affiliative motivation but a significant effect of confrontation type. Participants who received the racial confrontation were more likely to endorse positive stereotypes of Black people ($M = 3.95, SD = 0.86$) than those who received the rude confrontation ($M = 3.80, SD = 0.87$). There was no significant interaction between affiliative motivation and confrontation type on stereotype endorsement.

Intentions to Control Future Bias. Results from the two-way ANOVA with intentions to control future bias as the dependent variable found no main effect of affiliative motivation, but
a significant effect of confrontation type. Participants who received the racial confrontation were less likely to report intentions to monitor for bias ($M = 2.60$, $SD = 1.71$) than those who received a rude confrontation ($M = 2.91$, $SD = 1.70$). The interaction of affiliative motivation and confrontation type on intentions to control future bias was not significant.
Table 5. ANOVA results detailing the effects of affiliative motivation and confrontation type on all variables of interest.

<table>
<thead>
<tr>
<th>DV: Trust</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>1.15</td>
<td>0.94</td>
<td>.33</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>3.04</td>
<td>2.51</td>
<td>.11</td>
<td>.003</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>1.91</td>
<td>1.57</td>
<td>.21</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>1.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: NegSelf</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.57</td>
<td>0.58</td>
<td>.45</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>0.02</td>
<td>0.02</td>
<td>.89</td>
<td>.001</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>2.00</td>
<td>2.06</td>
<td>.15</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: NegOther</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.15</td>
<td>0.06</td>
<td>.81</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>25.20</td>
<td>9.78</td>
<td>.002**</td>
<td>.011</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>6.48</td>
<td>2.52</td>
<td>.11</td>
<td>.003</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>2.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Backlash</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.50</td>
<td>0.72</td>
<td>.40</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>4.54</td>
<td>6.52</td>
<td>.01*</td>
<td>.007</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.05</td>
<td>0.07</td>
<td>.79</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Attitudes Toward Black People</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>3.27</td>
<td>1.38</td>
<td>.24</td>
<td>.002</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>7.00</td>
<td>2.96</td>
<td>.09</td>
<td>.003</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>.99</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>2.36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Stereotype Endorsement</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
<td>.91</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>3.62</td>
<td>4.87</td>
<td>.03*</td>
<td>.006</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.18</td>
<td>0.24</td>
<td>.63</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Intentions to Control Future Bias</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>3.21</td>
<td>1.10</td>
<td>.29</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>15.86</td>
<td>5.44</td>
<td>.02*</td>
<td>.006</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>1.72</td>
<td>0.59</td>
<td>.44</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>864</td>
<td>2.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: MS = Mean squares, Effect size $= \eta^2_p$, *p < .05, **p < .01
Moderated Mediation

To test hypotheses 1b-4, I separated cases by confrontation type (racial v. rude) and used the PROCESS macro for SPSS which estimates a process model for moderated mediation (Model 8; Preacher et al., 2007) with bootstrapping \((n = 5000)\). Affiliative motivation (low = 0 v. high = 1) served as the independent variable, NegSelf and NegOther were mediators, the combined trust scale served as the moderator, and backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias were tested separately as the dependent variables. The indices for moderated mediation for all models regardless of dependent variables were not significant (see Table 6). The following results for each hypothesis are taken from pieces of the overall models.²

Table 6. PROCESS Model 8 results depicting the index of moderated mediation for the full model with affiliative motivation serving as the independent variable, negative self- and other-directed emotions acting as mediators, and backlash and prejudice reduction outcomes as dependent variables.

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>NegSelf</th>
<th>NegOther</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index (SE)</td>
<td>LLCI, ULCI</td>
</tr>
<tr>
<td>Backlash</td>
<td>-0.001 (0.01)</td>
<td>[-0.02, 0.02]</td>
</tr>
<tr>
<td>Attitudes (Black People)</td>
<td>0.001 (0.02)</td>
<td>[-0.05, 0.04]</td>
</tr>
<tr>
<td>Stereotype Endorsement</td>
<td>0.001 (0.02)</td>
<td>[-0.03, 0.03]</td>
</tr>
<tr>
<td>Intentions to Control Bias</td>
<td>-0.003 (0.06)</td>
<td>[-0.13, 0.11]</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td>NegSelf</td>
<td>NegOther</td>
</tr>
<tr>
<td></td>
<td>Index (SE)</td>
<td>LLCI, ULCI</td>
</tr>
<tr>
<td>Backlash</td>
<td>0.002 (0.01)</td>
<td>[-0.01, 0.02]</td>
</tr>
<tr>
<td>Attitudes (Black People)</td>
<td>0.001 (0.01)</td>
<td>[-0.02, 0.02]</td>
</tr>
<tr>
<td>Stereotype Endorsement</td>
<td>-0.002 (0.01)</td>
<td>[-0.02, 0.01]</td>
</tr>
<tr>
<td>Intentions to Control Bias</td>
<td>0.01 (0.03)</td>
<td>[-0.06, 0.08]</td>
</tr>
</tbody>
</table>

2. See Appendix X for more information about the results of these analyses when I adopted a more stringent data screening process.

Hypothesis 1b. I examined the highest order unconditional interactions of affiliative motivation and trust on negative self- and other-directed emotions to see whether trust moderated
the strength of the association between affiliative motivation with the confronter and emotional reactions after a racial (v. rude) confrontation. Results did not support hypothesis 1, finding that trust did not moderate the association between affiliative motivation and negative self-directed emotions, $F(1, 405) = 0.004, p = .95, R^2 = 0.001$, or other-directed emotions, $F(1, 405) = 2.73, p = .10, R^2 = 0.01$, in the racial confrontation condition. Likewise, the same results occurred for negative self-directed emotions, $F(1, 456) = 0.13, p = .72, R^2 = 0.001$, and other-directed emotions, $F(1, 456) = 0.05, p = .83, R^2 = 0.001$, in the rude confrontation condition.

The same analysis using only the outgroup trust measure rather than the combined trust measure indicated a marginally significant effect, such that outgroup trust moderated the association between affiliative motivation and negative other-directed affect for the racial, $F(1, 405) = 3.72, p = .05, R^2 = 0.01$, but not rude, $F(1, 456) = 0.00, p = .99, R^2 = 0.001$, confrontation condition. For participants in the low affiliative motivation condition, as trust increased ($b = -0.50, 95\% \text{ CI } [-0.94 \text{ – } -0.05]$), negative other-directed emotions decreased. Trust did not impact the association between high affiliative motivation and negative other-directed emotions.

**Hypothesis 2.** I sought to replicate past research and test whether negative emotionality mediated the relationship between confrontation content and the outcomes of backlash and prejudice reduction. I predicted that for participants confronted for racism (v. rudely), as negative self-directed emotions increase, backlash and prejudice will decrease and as negative other-directed emotions increase, backlash and prejudice will increase. I examined the beta coefficients and 95% confidence intervals of the model to draw conclusions about significance.

**Backlash.** Partially supporting hypothesis 2 for negative self-directed affect, the model indicated that as negative self-directed emotions increased, backlash decreased. However, the effect was significant for both the racial ($b = 0.09, 95\% \text{ CI } [0.04 \text{ – } 0.15]$) and rude ($b = 0.08,$
95% CI [0.03 – 0.13]) confrontation conditions. Confirming hypothesis 2 for negative other-directed affect, the results showed that as negative other-directed emotions increased, backlash increased for the racial ($b = -0.04, 95\% \text{ CI } [-0.08 \text{ – } -0.00]$), but not rude ($b = -0.03, 95\% \text{ CI } [-0.06 \text{ – } 0.01]$) confrontation condition.

**Attitudes Toward Black People.** Contrary to hypothesis 2, as negative self-directed emotions increased, positive attitudes toward Black people decreased in the racial ($b = -0.21, 95\% \text{ CI } [-0.37 \text{ – } -0.05]$), compared to rude ($b = 0.05, 95\% \text{ CI } [-0.12 \text{ – } 0.21]$) confrontation condition – suggesting increased prejudice. A similar pattern emerged for attitudes toward other groups of color (i.e., Mexican and Indigenous people). There was no effect of negative other-directed emotions on attitudes toward Black people in either confrontation condition.

**Stereotype Endorsement.** Mirroring the results for attitudes toward Black people, as negative self-directed emotions increased, positive stereotype endorsement decreased ($b = -0.16, 95\% \text{ CI } [-0.25 \text{ – } -0.08]$) in the racial versus rude ($b = -0.07, 95\% \text{ CI } [-0.16 \text{ – } 0.02]$) confrontation condition. Again, this effect suggests an increase in prejudice rather than the predicted reduction. There was no effect of negative other-directed emotions on stereotype endorsement regardless of confrontation type.

**Intentions to Control Future Bias.** Results partially supported hypothesis 2, such that as negative self-directed emotions increased, intentions to control future bias increased. However, the effect was significant for both the racial ($b = 0.61, 95\% \text{ CI } [0.45 \text{ – } 0.76]$) and rude ($b = 0.13, 95\% \text{ CI } [0.22 \text{ – } 0.56]$) confrontation conditions. Contrary to hypothesis 2, as negative other-directed emotions increased, intentions to control future bias decreased in the rude ($b = 0.14, 95\% \text{ CI } [0.02 \text{ – } 0.25]$), but not racial ($b = 0.06, 95\% \text{ CI } [-0.04 \text{ – } 0.15]$) confrontation condition.
**Hypothesis 3.** The model also tested whether negative emotionality mediated the association between affiliative motivation with the confronter and backlash and prejudice reduction outcomes following a racial (v. rude) confrontation. I examined the model’s indirect effects of negative self- and other-directed emotions on the link between affiliative motivation and backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias and their 95% confidence intervals to draw conclusions regarding significance.

The association between affiliative motivation and negative self-directed emotions was not significant for either the backlash or prejudice reduction outcomes for either the racial \( b = -0.15, 95\% \text{ CI } [-0.75 - 0.45] \) or rude \( b = 0.00, 95\% \text{ CI } [-0.54 - 0.55] \) confrontation conditions. Likewise, the same null effects occurred for affiliative motivation predicting negative other-directed emotions for the racial \( b = 0.52, 95\% \text{ CI } [-0.40 - 1.44] \) and rude \( b = -0.001, 95\% \text{ CI } [-0.82 - 0.82] \) confrontations. Therefore, hypothesis 3 was not supported as there was no viable link between affiliative motivation and negative emotionality, causing the mediation chain to break down for all dependent variables of interest.

**Hypothesis 4.** The model tested the hypothesis that trust moderated the strength of the association between affiliative motivation with the confronter and backlash and prejudice reduction outcomes after a racial (v. rude) confrontation. To determine significance, I examined the highest order unconditional interactions of affiliative motivation and trust on backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias.

The results indicated that trust did not moderate the association between affiliative motivation and backlash for either the racial, \( F (1, 403) = 2.74, p = .10, R^2 = 0.002 \), or rude, \( F (1, 454) = 0.03, p = .87, R^2 = 0.001 \), confrontation conditions. The same analysis using only the outgroup trust measure rather than the combined trust measure showed that outgroup trust
moderated the association between affiliative motivation and backlash for the racial, $F(1, 403) = 4.75, p = .03, R^2 = 0.004$, but not rude, $F(1, 454) = 0.05, p = .82, R^2 = 0.001$, confrontation condition. For participants in the low affiliative motivation condition, as outgroup trust increased ($b = -0.18, 95\% \text{ CI } [-0.34 \text{ – } -0.02]$) liking for the confronter increased above that reported by participants in the high affiliative motivation condition.

In terms of prejudice reduction, trust did not moderate the association between affiliative motivation and attitudes toward black people (racial: $F(1, 403) = 0.03, p = .85, R^2 = 0.001$; rude: $F(1, 454) = 0.59, p = .44, R^2 = 0.001$), or stereotype endorsement (racial: $F(1, 403) = 1.52, p = .22, R^2 = 0.003$; rude: $F(1, 454) = 2.52, p = .11, R^2 = 0.01$) in either the racial or rude confrontation condition.

Trust moderated the association between affiliative motivation and intentions to control future bias at high levels of trust ($b = -0.53, 95\% \text{ CI } [-0.96 \text{ – } -0.09]$) for the racial, $F(1, 403) = 6.88, p = .01, R^2 = 0.01$, but not rude, $F(1, 454) = 2.05, p = .15, R^2 = 0.004$, confrontation. In other words, this finding suggests that for participants who felt low affiliative motivation, as trust increased intentions to control future bias increased. Trust did not impact the association between affiliative motivation and intentions to control future bias for those with high affiliative motivation (see Table 7 and Figure 5).
Table 7. PROCESS Model 8 results depicting the interaction of affiliative motivation and trust on intentions to control future bias after a racial confrontation.

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Intentions to Control Future Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
</tr>
<tr>
<td>Affiliative Motivation (AM)</td>
<td>1.00 (0.45)*</td>
</tr>
<tr>
<td>NegSelf</td>
<td>0.61 (0.08)**</td>
</tr>
<tr>
<td>NegOther</td>
<td>0.06 (0.12)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.60 (0.12)**</td>
</tr>
<tr>
<td>AM x Trust</td>
<td>-0.37 (0.14)**</td>
</tr>
</tbody>
</table>

**Conditional effects of AM x Trust on Intentions to Control Future Bias at three levels of trust.** Analyses use continuous variables; effects below are one standard deviation +/- the mean.

<table>
<thead>
<tr>
<th>Trust</th>
<th>b (SE)</th>
<th>[LLCI, ULCI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ($M = 1.75$)</td>
<td>0.35</td>
<td>(-0.11, 1.82)</td>
</tr>
<tr>
<td>Medium ($M = 3.03$)</td>
<td>-0.12</td>
<td>(-0.42, 0.19)</td>
</tr>
<tr>
<td>High ($M = 4.14$)</td>
<td>-0.53</td>
<td>(-0.96, -0.09)</td>
</tr>
</tbody>
</table>

*Note: *p < .05, **p < .01

Figure 5. Plot depicting trust moderating the association between affiliative motivation and intentions to control future bias after a racial confrontation.
Study 2 Discussion

The results of Study 2 partially replicated those of Study 1. Trust, once again, influenced the association between affiliative motivation and negative other-directed emotions, but not negative self-directed emotions (hypothesis 1b). As in Study 1, trust was not correlated with affiliative motivation, but it did mitigate negative other-directed emotions for those with low affiliative motivation. As such, for people who were low in the desire to get along with their confronter, as trust for their confronter increased, negative other-directed emotions, like anger and annoyance, decreased regardless of the content of the confrontation. The impact of trust on negative other-directed emotions for those high in the desire to get along with their confronter was not significant, and overall levels of negative other-directed emotions were low. Thus, both studies consistently laid the groundwork for identifying trust as an important concept to consider in the outcomes of confrontation.

Importantly, this effect was marginally significant for outgroup trust only, rather than the combined, outgroup and workplace trust scale as in Study 1. One reason for this result might point to the assumptions of each scale. The items in the workplace trust scale (Robinson, 1996) leaned heavily upon one’s evaluation of another’s integrity (i.e., openness and honesty), while the outgroup trust scale (Montoya & Pittinsky, 2011) relied on evaluations of intentional benevolence. Study 2’s affiliative motivation manipulation included a level of skepticism on the part of the interaction partner about the participant’s existence, which may have impacted the way participants perceived their partner’s openness to the experience (i.e., workplace trust), leaving perceived intentions of benevolence intact (i.e., outgroup trust).

Study 2 also explored the impact of negative emotionality on backlash and prejudice reduction outcomes (hypothesis 2). Consistent with current predictions and prior work (Czopp et
al., 2006), as participants’ negative other-directed emotions increased, backlash increased for only those confronted for racism. On the other hand, as negative self-directed emotions, like guilt and self-criticism, increased after either confrontation (i.e., racial or rude), backlash decreased. Although research has yet to be published about the interplay between negative self- and other-directed affect following a confrontation, these results only partially support this study’s prediction that as negative self-directed emotions increase, backlash would decrease after a racial (v. rude) confrontation. The fact that the effect appeared in both the racial and rude confrontation conditions for negative self-directed affect could have been due to participants’ altogether low levels of reporting negative self-directed emotions in this study.

Negative emotionality had several interesting effects on prejudice reduction in Study 2 (hypothesis 2). Participants who felt increased negative self-directed emotions following a confrontation reported increased intentions to control future bias. Consistent with previous research, as people felt more guilt following an act of prejudice (Devine et al., 1991), they conveyed more interest in disrupting their bias and preventing future expressions of prejudice (Monteith et al., 1993). Contrary to predictions, the effect of negative self-directed emotions on intentions to control future bias occurred for participants confronted for racism and rudely. With that, it is important to consider the time in history during which these studies were conducted – early 2021 – amidst a racial reckoning in the United States. White people were, for the first time, overwhelmingly acknowledging that anti-Black racism is an issue (Onwuachi-Willig, 2021). Many of the White participants in these studies who were confronted rudely may have been exceptionally more willing to report their likelihood of monitoring for future bias and racism simply because they recognized a cultural shift in views on racism and it was at the forefront of
their minds. The data do not suggest a ceiling effect, so further results including the measure of intentions to control future bias exhibit variability and remain interpretable.

Unfortunately, negative self-directed emotions had the unintended effect of decreased positive stereotype endorsement and increased negative attitudes toward Black people in this study. As participants confronted for racism (v. rudely) reported increased emotions like guilt and self-criticism, they expressed more dislike for Black people and other groups of color as well as more endorsement of Black people as violent, lazy, and irresponsible. Again, considering the renewed focus on racism during the time this study was executed, feeling negatively about oneself because of bias might have backfired. Those individuals who were motivated to monitor their future outward bias might have felt external pressure to do so, but research has shown that external motivation to control prejudice often translates to behavioral, but not attitudinal change (Plant & Devine, 1998). White participants confronted for racism during this study might not have been able or willing to successfully suppress their prejudicial attitudes, resulting in greater reporting of prejudice (Hausmann & Ryan, 2004). Future research should include measures of internal and external motivation to control prejudice to potentially address these unpredicted effects.

Study 2 did not find support for negative emotionality mediating the association between affiliative motivation and backlash or prejudice reduction outcomes (hypothesis 3). Prior work has found that close others have an increased likelihood of experiencing guilt after conflict (Baumeister et al., 1994), and a decreased chance of enacting negative, retaliatory behaviors toward partners with whom they share emotional capital (Feeney & Lemay, 2012). However, affiliative motivation did not predict negative self- or other-directed emotions in this study, rendering the mediation chain inviable.
Although attempts were made to strengthen the affiliative motivation manipulation in Study 2, the online environment proved difficult once again. Unlike study 1, participants in the high affiliative motivation condition felt no more connected with their interaction partner than those in the low affiliative motivation condition. There was variability in the manipulation, such that excluding participants based on the affiliative motivation manipulation check decreased the sample by almost half – negatively impacting power to detect effects and negating the benefits of random assignment. Associations including affiliative motivation that emerged when applying that exclusion should be interpreted with caution and must be replicated. My previous unpublished work (Patrianakos, 2018) found that an affiliative motivation manipulation, like real friendship bonds, protected confronters against backlash after a racial confrontation. However, the online affiliative motivation manipulation was simply not as effective as an in-person manipulation has proven to be and these findings may not map onto actual relationship pairs.

Study 2 extended the work of Study 1 beyond trust’s influence on the association between affiliative motivation and negative emotionality to its impact on backlash and prejudice reduction (hypothesis 4). Trust influenced the association between affiliative motivation and backlash. For people low in motivation to get along with their confronter, as trust for their confronter increased, backlash decreased after a racial (v. rude) confrontation. The impact of trust on backlash for those high in motivation to get along with their confronter was not significant for either type of confrontation. Thus, trust had that same mitigating effect on backlash as it had in Study 1 for participants in the low affiliative motivation condition. This finding is consistent with past research that suggested close others who share emotional capital (Feeney & Lemay, 2012) and individuals with increased trust (Righetti et al., 2015; Campbell et al., 2010) are less likely to enact negative behaviors toward a partner after conflict. Like the
effect found in hypothesis 1b, this effect was only significant for the outgroup trust scale, rather than the combined outgroup and workplace scales. There was no strong prediction about whether one or both scales would be the best fit. In hindsight, the outgroup trust scale items accounted for the fact that there was no relational history between interaction partners, while the workplace trust scale may have relied too heavily on historical evidence of trust.

While trust did not influence the association between affiliative motivation and attitudes toward Black people or stereotype endorsement, it did impact intentions to control future bias. As predicted, for participants who were low (v. high) in the desire to get along with an interaction partner who confronted them for racism (v. rudely), as trust increased, reported intentions to monitor future bias increased – again revealing trust as mitigating outward expressions of prejudice for people low in affiliative motivation. This result confirms and extends research showing that trust is associated with cooperation (Ross et al., 2014) and accommodation (Shallcross & Simpson, 2012) following interpersonal conflict.
CHAPTER FIVE

GENERAL DISCUSSION

Anti-Black racism persists in the United States (Dovidio et al., 2017), negatively impacting the daily lives of Black people (Stangor et al., 2003). White allies can play an important part in disrupting racism by confronting perpetrators. Confrontation works to reduce prejudice (Czopp et al., 2006) but is often coupled with social consequences, like backlash or being disliked (Kaiser & Miller, 2001; 2003). Some White people who are attempting to move the conversation about racial justice forward and contribute to the solution for this pressing social problem often fear the social consequences of confronting (Swim & Hyers, 1999). The purpose of the current research was to offer a unique contribution to the social psychological literature on the outcomes of confrontations of anti-Black racism by looking toward the influence of social bonds and trust. The studies attempted to understand whether White allies might be more amenable to using confrontation as a prejudice reduction strategy in their social circles if backlash can be minimized.

Harnessing the Power of Trust

In this research, trust proved to be an important mitigating factor for both negative emotions and backlash directed at confronters. I predicted that trust would moderate the association between affiliative motivation and negative emotionality. While there was no impact of trust on negative self-directed emotions, trust did play a mitigating role for negative other-directed emotions. For people who were low in the desire to get along with their confronter,
increased trust reduced negative other-directed emotions, like anger and annoyance. In fact, as trust increased, negative other-directed emotions expressed by those with low affiliative motivation decreased to levels similar to those reported by people with high affiliative motivation. This finding replicates past research suggesting that as trust increases, negative emotions like contempt decrease (Kim et al., 2015). However, the results further extend the literature to reveal the strength of trust as mitigating negative emotionality, especially for those who do not share a social bond. Although these results were true for both racial and rude confrontations, they provide an initial glimpse into the power of harnessing trust in interpersonal conflict about racism.

Because prior research has found that increased negative other-directed emotions are often linked to increased backlash (Monteith et al., 1993; Czopp et al., 2006), this research sought to examine the impact of trust on backlash. Just as trust mitigated negative other-directed emotions, it extended its reach to mitigate backlash in the current work as well. Trust moderated the association between affiliative motivation and backlash, such that for people low in the desire to get along with their confronter, as trust increased, backlash decreased. I conducted a series of related studies in which high affiliative motivation and friendship protected confronters against backlash after a racial confrontation (Patrianakos, 2018), and these results mirror those findings while providing a potential explanatory mechanism for the effect – trust. In the current studies, as trust increased for the people low in affiliative motivation, their expressions of backlash decreased to levels reported by those with high affiliative motivation. This effect was significant for the racial, but not rude, confrontation conditions suggesting that trust plays a particularly important role in backlash outcomes for confrontations of anti-Black racism rather than general interpersonal conflict.
It is important to note that affiliative motivation and trust were not correlated in either of the present studies. This finding demonstrates that trust is not merely the motivation to affiliate, but its own conceptual variable independent of a shared social bond between two people. As such, the results of this research align most closely with the view that trust can be formed spontaneously, upon one’s first impression with another (Huang & Murnighan, 2010; McKnight et al., 1998; Weber et al., 2004), rather than over time as personal history develops (Kramer, 1999). Thus, understanding the trust one possesses as yet another tool in the confrontation toolbox might persuade some White people to engage in confrontation over and above their personal history with the perpetrator.

**Replicating Effects of Negative Emotionality**

This research did not find the predicted novel impact of affiliative motivation on negative emotionality, but it did partially replicate existing literature in terms of negative emotionality’s influence on backlash. For example, consistent with prior work (Czopp et al., 2006), as participants’ negative other-directed emotions increased, backlash increased for people confronted for anti-Black racism rather than in a rude fashion. This finding further solidifies negative other-directed emotions’ impact on backlash during confrontations for racism.

As negative self-directed emotions, like guilt and self-criticism, increased for participants in the current research, backlash decreased. This general pattern of results is in line with previous research (Czopp et al., 2006). Also consistent with previous research (Devine et al., 1991; Monteith et al., 1993), as people in the current research felt more negative self-directed emotions, like guilt, they reported increased intentions to prevent future expressions of prejudice. Contrary to existing literature, the effect of negative self-directed emotions on backlash and intentions to control future bias occurred regardless of the type of confrontation participants
received. However, the overall pattern of findings is promising in that it hints at the role of negative self-directed emotions in the outcomes of confrontations.

Limitations and Future Directions

One possible reason for the null results of the hypotheses referencing affiliative motivation is the manipulation’s lack of effectiveness in an online environment. I attempted to strengthen the affiliative motivation manipulation before executing the second study by taking cues from past research (Huntsinger et al., 2016) and creating a more believable script. Unfortunately, there was still plenty of variability in the manipulation check for how close participants felt toward their interaction partner. Future research could explore more realistic methodologies for an online manipulation of affiliative motivation, such as a more interactive chat, or the promise of future interaction or collaboration with their partner. The studies could also be replicated with an in-person manipulation. Moreover, it would be interesting to investigate the results of the current studies using individuals with an existing relationship (i.e., friendship) to see if the effects replicate prior work (Patrianakos, 2018).

The time during which these studies were conducted must be taken into consideration, as this work could be impacted by current events. While the 21st century has seen its share of racial unrest, the White participants in these studies were living in a time when White people were generally increasing their support for the Black Lives Matter movement and overwhelmingly acknowledging, for the first time in history, the discrimination faced by Black Americans (Payne, 2021) in national polls. Thus, some of the null effects and unexpected results could be explained by this moment in time. For example, being flooded by media coverage of anti-Black racism, something that White people typically avoid speaking about (Hughes et al., 2006; Lesane-Brown, 2006), could explain the lack of differences seen in the racial and rude
confrontation conditions. The availability of information about a shift in cultural views of racism could have also contributed to increased feelings of external motivation to control bias. Feeling more external, rather than internal motivation to control prejudice, might be one reason for the unexpected effects of negative self-directed emotions eliciting more negative attitudes and stereotype endorsement of Black people. People may have been willing to change outward behaviors but not internal attitudes (Plant & Devine, 1998; Hausmann & Ryan, 2004).

For potential confronters to really use the lessons from this work, they must be able to induce trust without an existing relationship. Future research should look toward ways in which trust without personal history can be manipulated. For instance, signaling psychological safety through one’s own vulnerability is one way to induce trust (Wanless, 2016). White confronters may benefit from delivering a confrontation for anti-Black racism that is preceded by one’s own confession of self-consciousness when talking about race. This might allow the perpetrator the psychological safety of making a mistake without feeling outcast by the confronter – possibly changing attitudes without the associated backlash.

**Conclusion**

This work set out to test a potential interpersonal solution to anti-Black racism. One such solution was to identify the conditions under which White people might be more receptive to confrontational messages related to reducing the harm of racism for Black people, starting within their own social bonds. The current studies suggest that for White people with the desire to get along with and maintain a social bond with another, chances of backlash are lower than for those without a similar desire. Importantly, for those without the desire to get along with or maintain a social bond with a perpetrator of anti-Black racism, establishing trust may reduce backlash while still reducing future intentions for prejudiced behavior.
Should White allies come to engage in racial confrontation with close others and strangers alike, these interpersonal instances of bias confrontation could translate into the larger society. Just as interpersonal and institutional racism work hand in hand (Ture & Hamilton, 1968), interpersonal confrontations could impact institutional change. If White allies are less afraid of confronting their coworkers’ racist behavior, for example, systemic changes (e.g., policies, norms) within companies may result. Together, the current studies offer a solution to White allies losing their voice in racially charged conversations with loved ones, and at the most, inspire some White people to overcome barriers to confrontation and speak out for change regardless of their audience. Talking about anti-Black racism with White perpetrators can be seen as the first step in dismantling this persistent social problem that has taken the lives of so many.
APPENDIX A

ONLINE RECRUITMENT TEXT
Title: Reasoning Ability Study

Description: In this HIT you will connect with an interaction partner, complete a reasoning ability task, receive performance feedback, and report various thoughts, feelings and behaviors.

Criteria/Qualification Required: Must be age 18 and over, White, a United States resident and fluent in English.

Reward: $0.50 (possibility of earning an additional $0.50)

Time Allotted: 60 minutes

Keywords: research, psychology, survey, experiment, questionnaire, science

Survey Link: [link to the survey]

Please note: You must provide a participant code for the HIT to be approved
APPENDIX B

STUDY 1 INFORMED CONSENT
Title: Reasoning Ability Study

Researcher: Jamie Patrianakos

Introduction: You are being asked to take part in a research study being conducted by Jamie Patrianakos for a dissertation under the supervision of Dr. Robyn Mallett in the Department of Psychology at Loyola University of Chicago. Please read this form carefully and ask any questions you may have before deciding whether to participate in the study.

Purpose: You are invited to participate in research investigating individuals’ reasoning ability. Please know that you will not be informed of the full scope or hypotheses of the present study until after your participation.

Procedures: Participants will interact with another participant via computer mediated instant messaging, complete a reasoning ability activity observed by your partner, receive performance feedback, and answer questions about related thoughts, feelings, and behaviors.

Risks and Benefits: There are minimal risks that do not exceed a level that you may encounter during your normal daily activities. Your participation in this online study involves risks similar to a person’s everyday use of the Internet. There are no direct benefits to you for your participation, however if you have not participated in a psychological study before, this is a good opportunity to experience how psychological research is conducted.

Time Commitment: The experiment will take less than 30 minutes to complete.

Compensation: You will receive $0.50 for submitting this survey, with a $0.50 bonus for full completion and attention. The researcher reserves the right to deny payment if the study is not completed.

Confidentiality: Confidentiality will be maintained to the degree permitted by the technology used. Your individual privacy will be maintained in all published and written data from the study. Your name will not be connected to the information you provide, nor will your individual responses be identified in any research reports describing the study. As such, the researcher will be unable to extract anonymous data from the database should the participant wish it withdrawn. All information obtained during the study will remain confidential.

Joining of your own free will: Your participation is voluntary. You may withhold information that you do not wish to disclose, and you do not have to answer any questions that you do not wish to answer. You may choose not to serve as a participant or withdraw from this study at any time without penalty.

This study has been approved by the Loyola Institutional Review Board for the Protection of Human Subjects. If you have questions about your rights as a research participant, you may contact the Loyola University Office of Research Services at (773) 508-2689. If you have any questions about the study, please contact Jamie Patrianakos (email: jpatrianakos@luc.edu) or Dr. Mallett (phone: 773.508.3028 email: rmallett@luc.edu).

Participant Statement: I have read the explanation provided to me and I understand that by clicking the survey link, I am verifying that I am at least 18 years of age and that I voluntarily agree to participate in this study.
APPENDIX C

INTRODUCTION COVER STORY AND AVATAR SELECTION
Introduction to the Reasoning Ability Study

We are interested in how people judge the reasoning ability of others. In order to study this, you will be connected with another participant who is simultaneously completing a similar study using novel software currently undergoing beta-testing. Once connected, you will complete a reasoning ability task while the other participant observes your responses. After the task is complete, the other participant will give you feedback about your performance through an instant messaging feature contained in the software.

Please choose one of these randomly selected avatars that best represents you. This avatar will be visible to the other participant during the study, and their selection will be visible to you.

I understand that I will now be connected with another participant.

Confirm
APPENDIX D

AFFILIATIVE MOTIVATION PROMPT
Please wait while we connect you with another participant…

CONNECTED!

Please introduce yourself to your partner. Tell them your name, hometown, birthday, and favorite color.

Participant 1: [Free Response]

Participant 2 is typing…

(High affiliative motivation)

: Hi there…Wow! We have the same birthday! My name is Sam, my favorite color is gray, and I’m from a town called Argleton. How’s the weather by you?

(Low affiliative motivation)

: Hi there…My name is Sam, I’m from a town called Argleton, my birthday is Jan. 2nd, and my favorite color is gray. How’s the weather by you?

(All conditions)

Participant 1: [Free Response]

Participant 2 is typing…

Participant 2: Same here…It’s telling me to click a button to continue…

[Automatically move participant 1 to the next survey page]

(High affiliative motivation)

Thank you for your introductions! You both will be interacting for the remainder of the study, about 30 minutes.

(Low affiliative motivation)

Thank you for your introductions! You both will be interacting for about 5 minutes during the study.
APPENDIX E

AFFILIATIVE MOTIVATION CHECKS
Attention Checks

1. What, if anything, did you share in common with your partner?
   a. Hometown
   b. Birthday
   c. Favorite food
   d. I don’t remember
   e. None of the above

2. How long will you interact with your partner for this study
   a. 30 minutes
   b. 5 minutes
   c. I don’t remember

Manipulation Checks (1 not at all to 7 very much)

1. Please indicate how similar you feel to the other participant.
2. How alert do you feel right now?
3. Please indicate how close you feel to the other participant.
4. How many MTurk studies have you participated in? [Free Response]
APPENDIX F

TRUST MEASURES
Outgroup Trust Scale (Montoya & Pittinsky, 2011)

Instructions: Please indicate the extent to which you agree or disagree with the following statements when thinking about your interaction partner from 1 (not at all) to 7 (very much).

1. If given the opportunity, the other participant would probably exploit my trust in them. (REVERSE)
2. I believe that the other participant will look out for my interests.
3. During the interaction with the other participant, I believe that they acted benevolently.
4. If the other participant were placed in a situation where they could gain at my expense, I believe that they would do so. (REVERSE)
5. The other participant cannot be trusted. (REVERSE)
6. I believe that the other participant can do things that benefit me.
7. I believe that the interests of the other participant can benefit me.
8. The other participant has the ability to help me.

Workplace Trust Scale (Robinson, 1996)

Instructions: Please indicate the extent to which you agree or disagree with the following statements when thinking about your interaction partner from 1 (strongly disagree) to 7 (strongly agree).

1. I believe the other participant has high integrity.
2. I can expect the other participant to treat me in a consistent and predictable fashion.
3. The other participant is not always honest and truthful. (REVERSE)
4. Please select ‘Strongly Agree’ to this question.
5. In general, I believe the other participant’s motives and intentions are good.
6. I don’t think the other participant treats me fairly. (REVERSE)
7. The other participant is open and up front with me.
8. I’m not sure I fully trust the other participant. (REVERSE)
APPENDIX G

PHOTO-INFERENCE TASK
Photo-Inference Task (Czopp et al., 2006)

Instructions: You will now complete a reasoning ability task while the other participant observes your responses. When you are finished, they will provide feedback on your performance.

This reasoning task assesses your ability to describe people from a single photograph and just a small amount of information. You will be presented with a picture of a person, along with a sentence relevant to the person.

For example: “This person can be found in a theater.”

Your task is to generate an inference for the person, such as a job or a hobby the person is likely to have based on the photograph and the sentence. For example, you might generate MOVIE FAN because this guy looks like he could be a movie fan OR you might generate ACTOR because this man also looks like he could be an actor.

You should generate and type your inferences as quickly as possible, but make sure that you spend enough time so that you provide responses that reflect your reasoning about the photos and sentences. We generally find that people perform best when they give the first reasonable responses that come to mind.

SET 1

• This person can be found on the streets.

• This person uses needles for recreation.

• This person can be found behind bars.

• This person is often found in a school.
SET 1

• This person often speaks in front of large groups of people.

• This person is often behind a desk.

• This person works with food.

• This person helps people with their problems.

• This person often travels cross country.

• This person works with computers.

• This person can be found in a hospital.

• This person works with children.
SET 1
• This person works with numbers.

SET 1
• This person works with a team of others.

SET 1
• This person works outside.

SET 1
• This person works at home.

SET 1
• This person can be found at the mall.

SET 1
• This person wears a uniform to work.

SET 1
• This person is often on their phone.

SET 1
• This person is found behind the wheel of a car.
APPENDIX H

CONFRONTATION FEEDBACK
Please wait while the other participant provides feedback about your performance on the task.

(Racial confrontation)

: You should really try to think about Black people in other ways that are less prejudiced. It just seems that you are some kind of racist. You know what I mean?

(Rude confrontation)

: Were you even paying attention to what you were doing? It was impossible for me to follow you. You know what I mean?

Participant 1: [Free Response]

[Automatically move participant 1 to the next survey screen once response is recorded]

You will now complete the final set of questions. The other participant has moved on to another part of the survey and can no longer see your computer screen or responses.
APPENDIX I

NEGSELF & NEGOTHER QUESTIONNAIRE
NegSelf/NegOther (Czopp et al., 2006)

**Instructions:** Please rate the extent to which you feel the following emotions from 1 (*does not apply at all*) to 7 (*applies very much*).

1. Angry at myself (Self)
2. Please select ‘Applies Very Much’ to this question
3. Angry at others (Other)
4. Guilty (Self)
5. Irritated with others (Other)
6. Annoyed with myself (Self)
7. Disgusted with others (Other)
8. Disappointed with myself (Self)
9. Disgusted with myself (Self)
10. Regretful (Self)
11. Shameful (Self)
12. Self-critical (Self)
Demographic Items

1. Please indicate what you think the gender is of the other participant [Free Response]

2. Please indicate what you think the race is of the other participant [Free Response]

3. Please indicate your age: [Free Response]

4. Please indicate your gender:
   a. Man
   b. Woman
   c. Nonbinary/Genderqueer
   d. Another gender [Text Box]

5. Please indicate your race:
   a. American Indian or Alaska Native
   b. Asian
   c. White
   d. Black or African American
   e. Hispanic
   f. Latinx or Spanish Origin
   g. Native Hawaiian or Other Pacific Islander
   h. Some other race or origin [Text Box]
APPENDIX K

HYPOTHESIS GUESS & PERFORMANCE EVALUATION
Hypothesis Guess

1. If you had to guess, what would you say this study was trying to figure out? What was our hypothesis? [Free Response]
2. How much do you believe that another participant was connected with you during the study? (1 not at all to 7 very much)

Performance Evaluation

**Instructions:** Please indicate the extent to which you agree or disagree with the following statements about your performance in this study from 1 (strongly disagree) to 7 (strongly agree).

1. I took this study seriously.
2. You should use my data in your analyses because it is accurate.
APPENDIX L

STUDY 1 DEBRIEFING FORM
Debriefing

The present study investigates how perpetrators of discrimination respond when they are confronted for their behaviors. Most research finds that confronting prejudice is the best way to reduce its occurrence, but there has been very little research exploring how the relationship between the confronter and the perpetrator may play a role in the response to the confrontation. Given the applicability of trust to many aspects of social interactions, it is not far off to suggest that trust could play a role in the outcomes of racial confrontation as well. This study expects to build on current research to better understand how the relationship between perpetrators and confronters, and trust influences negative emotionality after a racial confrontation.

To test this question, we have first utilized a task that evokes stereotypical responding in order to create a discriminatory situation. Additionally, providing differing information regarding the time frame of interaction and specific demographic information varied the relationship between the confronter and the perpetrator. We wonder whether this inclination to get along with the other participant and feelings of trust will result in decreased negative emotions directed at the confronter and increased negative emotions directed toward oneself after the confrontation has occurred.

In reality, this “other participant” was a computer simulation giving responses according to a script to ensure that all participants have a similar experience in the study. The scripts were different for participants because we also wanted to test whether the type of confrontation received by the perpetrator would affect negative emotionality differently. Participants received either a message regarding racism or the speed by which they completed the task.

If you would like to learn more about the research that inspired the present studies, please contact Jamie Patrianakos, jpatrianakos@luc.edu, or Dr. Robyn Mallett, rmallett@luc.edu.

You may also wish to read the following articles:


APPENDIX M

STUDY 1 DATA SCREENING DETAILS
Table 8. Details of the data screening process for participant race (Step 1).

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>33 (34.7%)</td>
<td>27 (28.4%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>12 (12.6%)</td>
<td>23 (24.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>45 (47.4%)</td>
<td>50 (52.6%)</td>
<td>95 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rude Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>28 (36.4%)</td>
<td>26 (33.8%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>11 (14.3%)</td>
<td>12 (15.6%)</td>
</tr>
<tr>
<td>Total</td>
<td>39 (50.6%)</td>
<td>38 (49.4%)</td>
<td>77 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>61 (35.5%)</td>
<td>53 (30.8%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>23 (13.4%)</td>
<td>35 (20.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>84 (48.8%)</td>
<td>88 (51.2%)</td>
<td>172 (100%)</td>
</tr>
</tbody>
</table>

Table 9. Details of the data screening for adhering to the procedure (Step 3).

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>6 (37.5%)</td>
<td>4 (25.0%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4 (25.0%)</td>
<td>2 (12.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>10 (62.5%)</td>
<td>6 (37.5%)</td>
<td>16 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rude Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>7 (33.3%)</td>
<td>9 (42.9%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>3 (14.3%)</td>
<td>2 (9.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>10 (47.6%)</td>
<td>11 (52.4%)</td>
<td>21 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>13 (35.1%)</td>
<td>13 (35.1%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>7 (18.9%)</td>
<td>4 (10.8%)</td>
</tr>
<tr>
<td>Total</td>
<td>20 (54.1%)</td>
<td>17 (45.9%)</td>
<td>37 (100%)</td>
</tr>
</tbody>
</table>
APPENDIX N

STUDY 1 SUPPLEMENTARY RESULTS
Data Screening

The data for these supplemental analyses include one additional screening step (step 3). Of the total sample, 575 participants were White identified (step 1) and passed 50% or more of the 5 attention checks scattered throughout the task (step 2). In addition to the steps used to screen data for the main analysis, I applied an additional criterion to ensure that only data for whom the manipulation was effective was included in the analyses (i.e., responded at/above [high affiliative motivation] or at/below [low affiliative motivation] the scale mid-point of “How close do you feel to the other participant?”). With that, 370 participants passed the affiliative motivation manipulation check (step 3). Finally, 347 participants completed the procedure in good faith by indicating their birthday in the introduction – an essential component of the affiliative motivation manipulation (step 4; see Figure 6 and Tables 10, 11, and 12 for more details). These analyses included 191 fewer participants than the main analysis.
Figure 6. Details of the supplemental data screening process for Study 1.

**Data Screening Procedure**

*N (Attrition %)*

747 U.S Adults

- Identified as White

- Passed Attention Checks

- Passed Manipulation

- Passed Procedural Check

**172 Excluded**

- American Indian or Alaska Native (7); Asian (40); Black or African American (59); Hispanic (24); Multiracial (31); Latinx or Spanish Origin (2); Native Hawaiian or Other Pacific Islander (1); Another Race (8)

**0 Excluded**

- All participants passed 50% or more of attention checks

**205 Excluded**

- Participants responded at or above (high affiliative motivation) and at or below (low affiliative motivation) the scale midpoint on: “How close do you feel to the other participant?”

**23 Excluded**

- Participants did not indicate birthday in introduction
Table 10. Details of the supplemental data screening process for participant race (Step 1).

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>33 (34.7%)</td>
<td>27 (28.4%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>12 (12.6%)</td>
<td>23 (24.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>45 (47.4%)</td>
<td>50 (52.6%)</td>
<td>95 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rude Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>28 (36.4%)</td>
<td>26 (33.8%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>11 (14.3%)</td>
<td>12 (15.6%)</td>
</tr>
<tr>
<td>Total</td>
<td>39 (50.6%)</td>
<td>38 (49.4%)</td>
<td>77 (100%)</td>
</tr>
</tbody>
</table>

Table 11. Details of the supplemental data screening process for the affiliative motivation manipulation check (Step 3).

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>36 (37.5%)</td>
<td>33 (34.4%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>11 (11.5%)</td>
<td>16 (16.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>47 (49.0%)</td>
<td>49 (51.0%)</td>
<td>96 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rude Confrontation</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>36 (33.3%)</td>
<td>49 (45.0%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>12 (11.0%)</td>
<td>12 (11.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>48 (44.0%)</td>
<td>61 (56.0%)</td>
<td>109 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Trust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Total</td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
<td>72 (35.1%)</td>
<td>82 (40.0%)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>23 (11.2%)</td>
<td>28 (13.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>95 (46.3%)</td>
<td>110 (53.7%)</td>
<td>205 (100%)</td>
</tr>
</tbody>
</table>
Table 12. Details of the supplemental data screening process for adhering to the procedure (Step 4).

<table>
<thead>
<tr>
<th></th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
</tr>
<tr>
<td>Racial Confrontation</td>
<td></td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Total</td>
<td>8 (72.7%)</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td></td>
</tr>
<tr>
<td>Affiliative Motivation</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Total</td>
<td>7 (58.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>15 (65.2%)</td>
</tr>
</tbody>
</table>

**Manipulation Checks**

With these exclusions, there was still a strong correlation between participants’ feelings of closeness and similarity with their interaction partner, $r = 0.76$, $p < .001$ (see Table 13). Thus, the two variables were combined to create a variable representing the connection participants felt with their interaction partner. An independent samples t-test again confirmed, $t(345) = -17.14$, $p < .001$; $d = 1.00$, that participants in the high affiliative motivation condition ($M = 4.73$, $SD = 0.96$) felt more connected to their interaction partner compared to those in the low affiliative motivation condition ($M = 2.83$, $SD = 1.10$). In terms of participant alertness, an independent samples t-test, $t(345) = -2.21$, $p < .05$; $d = 0.24$) revealed that those in the high affiliative motivation condition ($M = 5.98$, $SD = 0.99$) were more alert than those in the low affiliative motivation condition ($M = 5.71$, $SD = 1.28$), though participants in both conditions were above the scale mid-point regarding alertness. There was again no difference in the number of studies participants in either condition have completed, $t(345) = -0.80$, $p = .42$; $d = 0.09$. 
Descriptive Statistics

After reviewing the bivariate correlations for all the variables of interest (see Table 13), I combined the two trust measures (i.e., outgroup trust and workplace trust) as they were highly correlated, $r = 0.86, p < .001$. Affiliative motivation and the overall trust measure were weakly correlated, $r = 0.11, p < .05$. Participants in the high ($M = 4.20, SD = 1.32$) affiliative motivation condition trusted their interaction partner more than those in the low affiliative motivation condition ($M = 3.92, SD = 1.22$), $t(345) = -2.08, p < .05; d = 0.22$.

Table 13. Supplemental bivariate correlation matrix detailing the associations between variables of interest.

<table>
<thead>
<tr>
<th>Measure (n = 347)</th>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affiliative Motivation</td>
<td>--</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Confrontation Type</td>
<td>--</td>
<td>.01</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Trust Condition</td>
<td>--</td>
<td>.02</td>
<td>.02</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Combined Trust Scale</td>
<td>4.06 (1.28)</td>
<td>.11*</td>
<td>-.08</td>
<td>.58**</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Outgroup Trust</td>
<td>3.91 (1.32)</td>
<td>.10</td>
<td>-.07</td>
<td>-.58**</td>
<td>.96**</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Workplace Trust</td>
<td>4.22 (1.34)</td>
<td>.11*</td>
<td>-.07</td>
<td>-.52*</td>
<td>.96**</td>
<td>.86**</td>
<td>_</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. NegSelf</td>
<td>1.74 (1.22)</td>
<td>-.14*</td>
<td>-.02</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
<td>.01</td>
<td>_</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NegOther</td>
<td>2.48 (1.68)</td>
<td>.03</td>
<td>.10</td>
<td>-.09</td>
<td>-.08</td>
<td>-.08</td>
<td>-.09</td>
<td>.33**</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>9. Closeness</td>
<td>3.74 (1.46)</td>
<td>.71**</td>
<td>.05</td>
<td>.01</td>
<td>.25**</td>
<td>.24**</td>
<td>.24**</td>
<td>.19**</td>
<td>.05</td>
<td>_</td>
</tr>
<tr>
<td>10. Similarity</td>
<td>3.85 (1.52)</td>
<td>.57**</td>
<td>.01</td>
<td>.05</td>
<td>.20**</td>
<td>.19**</td>
<td>.19**</td>
<td>.16**</td>
<td>.03</td>
<td>.76**</td>
</tr>
</tbody>
</table>

Note: $p < .05^*, p < .01^{**}$
Hypothesis 1A

First, I tested hypothesis 1a to determine whether trust measured before or after the racial (v. rude) confrontation moderated the strength of the association between affiliative motivation and negative self- and other-directed emotions. I started by conducting two three-way ANOVAs with affiliative motivation (high v. low), confrontation (racial v. rude), and trust time point (before v. after confrontation) as fixed factors, and NegSelf and NegOther, respectively, as the dependent variables. Any significant two- or three-way interactions including trust were examined to ascertain the time point driving the effect. Additionally, all analyses of variance controlled for participants’ evaluation of how seriously they took the study.

With NegSelf entered as the dependent variable, results with the additional exclusion criterion showed a main effect of affiliative motivation, such that participants who experienced high affiliative motivation ($M = 1.92, SD = 0.09$) felt more negative self-directed emotions than those in the low affiliative motivation condition ($M = 1.57, SD = 0.09$). The ANOVA again indicated no significant two- or three-way interactions between affiliative motivation and trust time point, confrontation type and trust time point, or affiliative motivation, confrontation type, and trust (see Table 14).

With the more rigorous exclusion criteria and NegOther entered as the dependent variable, results showed a main effect of confrontation type, such that participants who received the racial confrontation ($M = 2.67, SD = 0.13$) felt more negative other-directed emotions than those who received the rude confrontation ($M = 2.27, SD = 0.13$). Unlike the main analysis, the ANOVA indicated no significant two- or three-way interactions between affiliative motivation and trust time point, confrontation type and trust time point, or affiliative motivation, confrontation type, and trust time point (see Table 14).
Table 14. Supplemental ANOVA results detailing the effects of affiliative motivation, confrontation type, and trust time point on negative self- and other-directed emotions.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DV: NegSelf</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>10.82</td>
<td>7.59</td>
<td>.01*</td>
<td>.020</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>0.29</td>
<td>0.20</td>
<td>.89</td>
<td>.001</td>
</tr>
<tr>
<td>Trust (C)</td>
<td>1</td>
<td>0.25</td>
<td>0.17</td>
<td>.68</td>
<td>.001</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.16</td>
<td>0.11</td>
<td>.74</td>
<td>.001</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>0.63</td>
<td>0.44</td>
<td>.51</td>
<td>.001</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>0.61</td>
<td>0.43</td>
<td>.51</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>338</td>
<td>1.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DV: NegOther</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.79</td>
<td>0.29</td>
<td>.59</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>13.54</td>
<td>4.87</td>
<td>.03*</td>
<td>.014</td>
</tr>
<tr>
<td>Trust (C)</td>
<td>1</td>
<td>7.75</td>
<td>2.79</td>
<td>.10</td>
<td>.008</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>1.21</td>
<td>0.44</td>
<td>.51</td>
<td>.001</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>0.49</td>
<td>0.18</td>
<td>.68</td>
<td>.001</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>7.92</td>
<td>2.85</td>
<td>.09</td>
<td>.008</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>0.88</td>
<td>0.32</td>
<td>.57</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>338</td>
<td>2.78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05

**Hypothesis 1B**

The analyses of variance with this smaller sampler did not present a clear indication of whether trust measured before or after the racial (v. rude) confrontation moderated the association between affiliative motivation and negative self- and other-directed emotions (hypothesis 1a). I next tested hypothesis 1a and 1b (i.e., when affiliative motivation was high, as trust for the confronter increased, negative self-directed affect increased, but there was no impact on negative other-directed affect; when affiliative motivation was low, as trust for the confronter increased, negative other-directed affect decreased, but there was no impact of trust on negative self-directed affect) simultaneously. I used the PROCESS macro for SPSS which estimates a
process model for moderation (Model 2; Preacher et al., 2007) with bootstrapping ($n = 5000$). I separated the cases by trust time point (before v. after the confrontation) to examine the results within each condition. In this analysis, affiliative motivation (low = 0 v. high = 1) served as the independent variable, confrontation type (racial = 1 v. rude = 0) and the continuous measure of trust were moderators, and negative self- and other-directed emotions served as dependent variables.

Similar to the main analysis, when trust was measured after the confrontation, the continuous trust variable did not moderate the association between affiliative motivation and negative self-directed emotions, $F(1, 155) = 0.80, p = .37, R^2 = .01$, in either the racial or rude confrontation condition, $F(1, 155) = 0.01, p = .92, R^2 = .00$. The results again showed a marginally significant effect such that trust measured after the confrontation moderated the association between affiliative motivation and negative other-directed emotions, $F(1, 155) = 3.79, p = .05, R^2 = .02$, regardless of confrontation type, $F(1, 155) = 0.19, p = .67, R^2 = .00$. However, when probed further, these analyses revealed that the moderation was driven by outgroup, $F(1, 155) = 4.25, p = .04, R^2 = .02$, as opposed to workplace trust, $F(1, 155) = 1.90, p = .17, R^2 = .00$. Meaning that regardless of whether participants are confronted for racism or rudely, when affiliative motivation is high, the impact of outgroup trust on negative other-directed affect is not significant but when affiliative motivation is low, as outgroup trust increases, negative other-directed affect decreases. These results partially support hypothesis 1b (see Table 15 and Figure 7).

When trust was measured before the confrontation the results mirrored the main analyses. The continuous trust variable did not moderate the association between affiliative motivation and negative self-directed emotions, $F(1, 150) = 0.76, p = .39, R^2 = .00$, in either the racial or rude
confrontation condition, \( F(1, 180) = 0.41, p = .52, R^2 = .00 \). Likewise, trust did not moderate the association between affiliative motivation and negative other-directed emotions, \( F(1, 180) = 0.60, p = .44, R^2 = .00 \), in either confrontation condition, \( F(1, 180) = 0.00, p = .99, R^2 = .00 \).

These results address hypothesis 1a such that trust measured after rather than before the confrontation moderates the association between affiliative motivation and negative emotionality.

Table 15. Supplemental PROCESS Model 2 results for Hypothesis 1a depicting the interaction between affiliative motivation and trust measured before and after a confrontation on negative self- and other-directed emotions.

<table>
<thead>
<tr>
<th>Before the Confrontation</th>
<th>Negative Self-Directed Emotions</th>
<th>Negative Other-Directed Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>[LLCI, ULCI]</td>
</tr>
<tr>
<td>Affiliative Motivation (AM)</td>
<td>-0.73 (1.09)</td>
<td>[-2.89, 1.42]</td>
</tr>
<tr>
<td>Confrontation</td>
<td>-0.10 (0.26)</td>
<td>[-0.61, 0.42]</td>
</tr>
<tr>
<td>AM x Confrontation</td>
<td>0.24 (0.37)</td>
<td>[-0.49, 0.96]</td>
</tr>
<tr>
<td>Trust (continuous)</td>
<td>-0.11 (0.16)</td>
<td>[-0.42, 0.21]</td>
</tr>
<tr>
<td>AM x Trust</td>
<td>0.19 (0.22)</td>
<td>[-0.25 – 0.63]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>After the Confrontation</th>
<th>Negative Self-Directed Emotions</th>
<th>Negative Other-Directed Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>[LLCI, ULCI]</td>
</tr>
<tr>
<td>Affiliative Motivation (AM)</td>
<td>-0.11 (0.60)</td>
<td>[-1.29, 1.07]</td>
</tr>
<tr>
<td>Confrontation</td>
<td>-0.17 (0.28)</td>
<td>[-0.72, 0.38]</td>
</tr>
<tr>
<td>AM x Confrontation</td>
<td>0.04 (0.39)</td>
<td>[-0.72, 0.80]</td>
</tr>
<tr>
<td>Trust (continuous)</td>
<td>-0.01 (0.12)</td>
<td>[-0.24, 0.22]</td>
</tr>
<tr>
<td>AM x Trust</td>
<td>0.14 (0.16)</td>
<td>[-0.17 – 0.45]</td>
</tr>
</tbody>
</table>

Conditional effects of AM x Trust measured after the confrontation on NegOther at three levels of trust collapsed across confrontation condition. Analyses use continuous variables; effects below are one standard deviation +/- the mean.

<table>
<thead>
<tr>
<th>Trust</th>
<th>b (SE)</th>
<th>[LLCI, ULCI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (M = 1.71)</td>
<td>-0.42 (0.37)</td>
<td>[-1.16, 0.32]</td>
</tr>
<tr>
<td>Medium (M = 3.38)</td>
<td>0.25 (0.23)</td>
<td>[-0.22, 0.71]</td>
</tr>
<tr>
<td>High (M = 4.34)</td>
<td>0.63 (0.31)*</td>
<td>[0.02, 1.24]</td>
</tr>
</tbody>
</table>
After excluding the additional 191 participants for not passing the manipulation check, the supplemental analyses revealed some notable differences compared to the main analyses. Participants in the high affiliative motivation condition reported being more alert than those in the low affiliative motivation condition. Affiliative motivation and trust were correlated, showing that participants in the high affiliative motivation condition trust their interaction partner more than those in the low affiliative motivation condition. The ANOVA found additional main effects of affiliative motivation on negative self-directed emotions and confrontation type on negative other-directed emotions. However, there was no interaction of confrontation type and trust time point, making the interpretation of the results for hypothesis 1a
difficult. The results of the moderation model mirrored the main analysis but specified that outgroup trust was driving the effect.

There are several reasons the sample with these more rigorous exclusions was not chosen as the optimal choice compared to the main analysis. The number of exclusions from the sample places the power to detect effects at risk. Random assignment was violated in this sample, making causal interpretations of the results difficult. Additionally, the correlation between affiliative motivation and trust suggested that the two were associated using these data, which could have been an artifact of the violation of random assignment. Overall, the affiliative motivation manipulation was not the strongest given these and the main analyses.
APPENDIX O

STUDY 2 INFORMED CONSENT
Consent to Participate in Research Project

Title: Reasoning Ability Study  
Researcher: Jamie Patrianakos

Introduction: You are being asked to take part in a research study being conducted by Jamie Patrianakos for a dissertation under the supervision of Dr. Robyn Mallett in the Department of Psychology at Loyola University of Chicago. Please read this form carefully and ask any questions you may have before deciding whether to participate in the study.

Purpose: You are invited to participate in research investigating individuals’ reasoning ability. Please know that you will not be informed of the full scope or hypotheses of the present study until after your participation.

Procedures: Participants will interact with another participant via computer mediated instant messaging, complete a reasoning ability activity observed by your partner, receive performance feedback, and answer questions about related thoughts, feelings, and behaviors.

Risks and Benefits: There are minimal risks that do not exceed a level that you may encounter during your normal daily activities. Your participation in this online study involves risks similar to a person’s everyday use of the Internet. There are no direct benefits to you for your participation, however if you have not participated in a psychological study before, this is a good opportunity to experience how psychological research is conducted.

Time Commitment: The experiment will take less than 40 minutes to complete.

Compensation: You will receive $0.50 for submitting this survey, with a $0.50 bonus for full completion and attention. The researcher reserves the right to deny payment if the study is not completed.

Confidentiality: Confidentiality will be maintained to the degree permitted by the technology used. Your individual privacy will be maintained in all published and written data from the study. Your name will not be connected to the information you provide, nor will your individual responses be identified in any research reports describing the study. As such, the researcher will be unable to extract anonymous data from the database should the participant wish it withdrawn. All information obtained during the study will remain confidential.

Joining of your own free will: Your participation is voluntary. You may withhold information that you do not wish to disclose, and you do not have to answer any questions that you do not wish to answer. You may choose not to serve as a participant or withdraw from this study at any time without penalty.

This study has been approved by the Loyola Institutional Review Board for the Protection of Human Subjects. If you have questions about your rights as a research participant, you may contact the Loyola University Office of Research Services at (773) 508-2689. If you have any questions about the study, please contact Jamie Patrianakos (email: jpatrianakos@luc.edu) or Dr. Mallett (phone: 773.508.3028 email: rmallett@luc.edu).

Participant Statement: I have read the explanation provided to me and I understand that by clicking the survey link, I am verifying that I am at least 18 years of age and that I voluntarily agree to participate in this study.
APPENDIX P

STUDY 2 INTRODUCTION ADDITIONS
You will now introduce yourself to your interaction partner.

Please select any number below. A computer algorithm will use this number to randomly select one person to choose the introduction topics and the other to start the introduction first.

[1-10]

You selected [Participant’s number choice]

The computer’s algorithm randomly chose you to introduce yourself first and the other participant to choose 3 introduction topics from this list:

Hometown        Relationship Status
Dream Car        Career
Birthday         Hobbies
Current Location Eye Color
Favorite Color   Favorite Food

The button to continue will appear after a few seconds, leaving you time to read the full list of topics.

Participant 2 is selecting topics…

Participant 2 has chosen the introduction topics.

Hometown
Birthday
Favorite Color

Note: This information will not be attached to your data. It will only be used for introductory purposes.

Please introduce yourself to your interaction partner by telling them your first name, hometown, birthday (month and day ONLY), and favorite color.
APPENDIX Q

STUDY 2 CONFRONTATION FEEDBACK ADDITIONS
(High affiliative motivation)

:Hi there…Wow! We have the same birthday! My name is Sam, my favorite color is grey, and I’m from a town called Argleton. I wonder if I’m really chatting with someone…Am I? Can you tell me something unique, like how’s the weather by you?

(Low affiliative motivation)

:Hi there…My name is Sam, I’m from a town called Argleton, my birthday is Jan. 2\textsuperscript{nd}, and my favorite color is grey. I wonder if I’m really chatting with someone…Am I? Can you tell me something unique, like how’s the weather by you?
APPENDIX R

STUDY 2 DEBRIEFING FORM
Debriefing

The present study investigates how perpetrators of discrimination respond when they are confronted for their behaviors. Most research finds that confronting prejudice is the best way to reduce its occurrence, but there has been very little research exploring how the relationship between the confronter and the perpetrator may play a role in the response to the confrontation. Given the applicability of trust to many aspects of social interactions, it is not far off to suggest that trust could play a role in the outcomes of racial confrontation as well. This study expects to build on current research to better understand how the relationship between perpetrators and confronters, and trust influences negative emotionality, backlash, and prejudice reduction after a racial confrontation.

To test this question, we have first utilized a task that evokes stereotypical responding in order to create a discriminatory situation. Additionally, providing differing information regarding the time frame of interaction and specific demographic information varied the relationship between the confronter and the perpetrator. Research has shown that individuals who feel a stronger motivation to get along with another person will “tune” their thoughts and behavior to be in-line with the other person’s. Those individuals who are low in this motivation to get along will not experience the same effect. The hope is that by providing a context in which participants feel the motivation to get along with the other participant (i.e., more time interacting and shared demographics), the more they will change their discriminatory behavior after the task and confrontation to be more egalitarian. Moreover, we wonder whether this inclination to get along with the other participant and feelings of trust will result in changes in negative emotionality and decreases in backlash and prejudice.

In reality, this “other participant” was a computer simulation giving responses according to a script to ensure that all participants have a similar experience in the study. The scripts were different for participants because we also wanted to test whether the type of confrontation received by the perpetrator would affect prejudice reduction and social consequences differently. Participants received either a message regarding racism or the speed by which they completed the task.

If you would like to learn more about the research that inspired the present studies, please contact Jamie Patrianakos, jpatrianakos@luc.edu, or Dr. Robyn Mallett, rmallett@luc.edu.

You may also wish to read the following articles:


APPENDIX S

LIKING MEASURE
Liking (Mallett & Wagner, 2011)

**Instructions:** Please indicate the extent to which you agree or disagree with the following statements about the other participant from 1 (*absolutely not*) to 5 (*absolutely*)

1. The other participant is too sensitive.
2. The other participant is likeable.
3. The other participant is good-natured.
4. Please select ‘Absolutely’ to this question.
5. The other participant is warm.
6. I have respect for the other participant.
7. I would want the other participant as a friend.
APPENDIX T

ATTITUDES MEASURE
Attitudes (Payne et al., 2010)

**Instructions:** Please indicate the extent to which you like or dislike each of the following groups from 1 (dislike a great deal) to 7 (like a great deal).

1. White people
2. Mexican people
3. Black people
4. Asian people
5. American Indian people
APPENDIX U

STEREOTYPE ENDORSEMENT MEASURE
Stereotype Endorsement (Payne et al., 2010)

**Instructions:** Please indicate how well each of the following words or phrases describe most Black people from 1 (*not well at all*) to 5 (*extremely well*).

1. Friendly
2. Determined to succeed
3. Law abiding
4. Hard-working
5. Intelligent at school
6. Smart at everyday things
7. Good neighbors
8. Dependable
9. Keep up their property
10. Violent
11. Boastful
12. Complaining
13. Lazy
14. Irresponsible
APPENDIX V

INTENTIONS TO CONTROL FUTURE BIAS MEASURE
Intentions to Control Future Bias (Parker et al., 2018)

Instructions: Please rate the extent to which you believe the following statements are true from 1 (not at all true) to 7 (completely true).

1. More so than before this study, I realize that I should be on guard so that race does not affect my judgments about people’s jobs or hobbies.
2. I am now more concerned that race might influence my personal evaluations of others after doing this study than I was before.
3. Please select ‘Completely True’ for this question.
4. If I were actually evaluating someone’s jobs or hobbies, I would now be more cautious about the possibility of race affecting my evaluation in a negative way.
5. I am now more concerned than I was before the study that I could evaluate a Black person less favorably than someone of another race because of the stereotypes associated with their racial group.
6. I am now more concerned than I was before participating in this study that I may judge the abilities of people from different racial or ethnic backgrounds unfairly.
7. After what I learned today, I will be more on guard for racially biased behavior.
8. I will now be more careful to monitor myself to make sure that I am not being racially biased.
9. Based on what I learned today, I have a better understanding of why Black people sometimes feel that they are not treated fairly.
10. I am concerned about exhibiting racial bias in the future.
11. I am now more concerned than I was before this study about biases based on race in our society.
APPENDIX W

STUDY 2 DATA SCREENING DETAILS
Table 16. Details of the data screening process for participant race (Step 1).

<table>
<thead>
<tr>
<th></th>
<th>High Affiliative Motivation</th>
<th>Low Affiliative Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racial Confrontation</td>
<td>112 (41.0%)</td>
<td>50 (18.3%)</td>
<td>162 (59.3%)</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td>74 (27.1%)</td>
<td>37 (13.6%)</td>
<td>111 (40.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>186 (68.1%)</td>
<td>87 (31.9%)</td>
<td>273 (100%)</td>
</tr>
</tbody>
</table>

Table 17. Details of the data screening process for adhering to the procedure (Step 3).

<table>
<thead>
<tr>
<th></th>
<th>High Affiliative Motivation</th>
<th>Low Affiliative Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racial Confrontation</td>
<td>7 (38.9%)</td>
<td>3 (16.7%)</td>
<td>10 (55.6%)</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td>5 (27.8%)</td>
<td>3 (16.7%)</td>
<td>8 (44.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>12 (66.7%)</td>
<td>6 (33.3%)</td>
<td>18 (100%)</td>
</tr>
</tbody>
</table>
APPENDIX X

STUDY 2 SUPPLEMENTARY RESULTS
Data Screening

The data for these supplemental analyses include one additional screening step (step 3). Of the total sample, 887 participants were White-identified (step 1) and passed 50% or more of the 7 attention checks scattered throughout the task (step 2). In addition to the steps used to screen data for the main analysis, I applied an additional criterion to ensure that only data for whom the manipulation was effective was included in the analyses (i.e., responded at/above [high affiliative motivation] or at/below [low affiliative motivation] the scale mid-point of “How close do you feel to the other participant?”). With that, 503 participants passed the affiliative motivation manipulation check (step 3). Finally, 497 participants completed the procedure in good faith by indicating their birthday in the introduction – an essential component of the affiliative motivation manipulation (step 4; see Figure 8 and Table 18, 19, and 20 for more details). These analyses included 372 fewer participants than the main analysis.
Figure 8. Details of the supplemental data screening process for Study 2.

Table 18. Details of the supplemental data screening process for participant race (Step 1).

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>High Affiliative Motivation</th>
<th>Low Affiliative Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>112 (41.0%)</td>
<td>50 (18.3%)</td>
<td>162 (59.3%)</td>
</tr>
</tbody>
</table>
Table 19. Details of the supplemental data screening process for the affiliative motivation manipulation check (Step 3).

<table>
<thead>
<tr>
<th></th>
<th>High Affiliative Motivation</th>
<th>Low Affiliative Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racial Confrontation</td>
<td>152 (39.6%)</td>
<td>24 (6.3%)</td>
<td>176 (45.8%)</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td>166 (43.2%)</td>
<td>42 (10.9%)</td>
<td>208 (54.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>318 (82.8%)</td>
<td>66 (17.2%)</td>
<td>384 (100%)</td>
</tr>
</tbody>
</table>

Table 20. Details of the supplemental data screening process for adhering to the procedure (Step 4).

<table>
<thead>
<tr>
<th></th>
<th>High Affiliative Motivation</th>
<th>Low Affiliative Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racial Confrontation</td>
<td>2 (33.3%)</td>
<td>2 (33.3%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>Rude Confrontation</td>
<td>1 (16.7%)</td>
<td>1 (16.7%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>3 (50.0%)</td>
<td>3 (50.0%)</td>
<td>6 (100%)</td>
</tr>
</tbody>
</table>

**Manipulation Checks**

With these exclusions \((n = 497)\), there was still a strong correlation between participants’ feelings of closeness and similarity with their interaction partner, \(r = 0.76, p < .001\) (see Table 21). Thus, the two variables were combined to create a variable representing the connection participants felt with their interaction partner. Unlike the main analysis, an independent samples t-test revealed, \(t(495) = -21.48, p < .001; d = 1.00\), that participants in the high affiliative motivation condition \((M = 4.66, SD = 0.84)\) felt more connected to their interaction partner compared to those in the low affiliative motivation condition \((M = 2.79, SD = 1.08)\).

Additionally, in terms of participant alertness, an independent sampled t-test, \(t(495) = -2.77, p < .01; d = 0.25\) revealed that those in the high affiliative motivation condition \((M = 6.04, SD = 0.99)\) were more alert than those in the low affiliative motivation condition \((M = 5.76, SD = \)
1.24), though participants in both conditions were above the scale mid-point regarding alertness. There was again no difference in the number of studies participants in either condition have completed, \( t(495) = -0.73, p = .47; d = .07 \).

**Descriptive Statistics**

The bivariate correlations for all the variables of interest (see Table 21), indicated that the two trust measures (i.e., outgroup trust and workplace trust) were again highly correlated with one another, \( r = 0.77, p < .001 \), so they were combined in subsequent analyses. Similar to the main analysis, the type of confrontation (i.e., racial or rude) was correlated with backlash, \( r = -0.10, p < .05 \), and an independent samples t-test confirmed, \( t(495) = 2.11, p < .05; d = .19 \), that those who experienced a racial confrontation (\( M = 2.28, SD = 0.85 \)) liked their confronter less than those confronted rudely (\( M = 2.44, SD = 0.81 \)). Confrontation type was correlated with stereotype endorsement in this smaller sample as well, \( r = 0.10, p < .05 \), and an independent samples t-test, \( t(495) = 2.32, p < .05; d = .21 \), found that those confronted for racism (\( M = 3.97, SD = 0.88 \)) reported more positive stereotype endorsement than those confronted rudely (\( M = 3.79, SD = 0.87 \)).

With the more rigorous exclusion criteria, affiliative motivation was significantly correlated with intentions to control future bias, \( r = 0.09, p < .05 \), and an independent samples t-test confirmed, \( t(495) = -2.03, p < .05; d = .18 \), that those in the high in affiliative motivation (\( M = 2.97, SD = 1.76 \)) reported greater intentions to control future bias than those in the low affiliative motivation condition (\( M = 2.66, SD = 1.68 \)). However, as in the main analysis, affiliative motivation and the combined trust measure were not correlated, \( r = 0.15, p = .31 \), and there was no difference between participants in the high or low affiliative motivation condition in terms of trusting their interaction partner, \( t(495) = -1.01, p = .31; d = .09 \).
Table 21. Supplemental bivariate correlation matrix detailing the associations between variables of interest.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affiliative Motivation</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Confront Type</td>
<td>-- .04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Closeness (1.46)</td>
<td>3.63</td>
<td>.72**</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Similarity (1.41)</td>
<td>3.81</td>
<td>.58**</td>
<td>.04</td>
<td>.76**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Combined Trust Scale (1.07)</td>
<td>3.11</td>
<td>.05</td>
<td>.04</td>
<td>.15**</td>
<td>.16**</td>
<td>.94**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Outgroup Trust (1.09)</td>
<td>2.98</td>
<td>.03</td>
<td>.06</td>
<td>.14**</td>
<td>.16**</td>
<td>.95**</td>
<td>.77**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Workplace Trust (1.18)</td>
<td>3.24</td>
<td>.06</td>
<td>.11</td>
<td>.13**</td>
<td>.12**</td>
<td>.12**</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NegSelf (0.99)</td>
<td>1.58</td>
<td>.04</td>
<td>.02</td>
<td>.11</td>
<td>.13**</td>
<td>.12**</td>
<td>.12**</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. NegOther (1.61)</td>
<td>2.55</td>
<td>.05</td>
<td>.07</td>
<td>.06</td>
<td>.05</td>
<td>-.42**</td>
<td>-.40**</td>
<td>-.39**</td>
<td>.29**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Backlash (0.83)</td>
<td>2.36</td>
<td>.08</td>
<td>-.10</td>
<td>.19**</td>
<td>.19**</td>
<td>.79**</td>
<td>.74**</td>
<td>.74**</td>
<td>.17**</td>
<td>-.36**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Attitudes Toward Black People (1.58)</td>
<td>5.54</td>
<td>.03</td>
<td>.08</td>
<td>.06</td>
<td>.06</td>
<td>.07</td>
<td>.04</td>
<td>-.12**</td>
<td>-.12**</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Stereotype Endorsement (0.88)</td>
<td>3.87</td>
<td>-.01</td>
<td>.10</td>
<td>.04</td>
<td>-.02</td>
<td>.08</td>
<td>.09</td>
<td>.07</td>
<td>-.16**</td>
<td>-.13**</td>
<td>.09</td>
<td>.64**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Intentions to Control Future Bias (1.73)</td>
<td>2.82</td>
<td>.09</td>
<td>-.10</td>
<td>.18**</td>
<td>.14**</td>
<td>.17**</td>
<td>.18**</td>
<td>.15**</td>
<td>.36**</td>
<td>.09</td>
<td>.21**</td>
<td>-.04</td>
<td>-.07</td>
<td></td>
</tr>
</tbody>
</table>

Note: p< .05*, p<.01**

Analyses of Variance

I began exploring patterns in the data that would later inform pieces of the full moderated mediation model (hypotheses 1b-4) by conducting several two-way ANOVAs. The fixed factors for each ANOVA were affiliative motivation (high v. low) and confrontation type (racial v. rude), and the dependent variables were trust, NegSelf, NegOther, backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias, respectively. Any significant two-way interactions were examined to determine whether the results were driven by the racial or rude confrontation condition. Additionally, all analyses controlled for participants’ evaluation of how seriously they took the study (see Table 22).
**Trust.** Similar to the results of the main analysis, the ANOVA indicated no significant main effects for affiliative motivation or confrontation type, nor a significant interaction.

**NegSelf.** The results showed no significant main effects of affiliative motivation or confrontation type on negative self-directed emotions. However, there was a significant two-way interaction using the more rigorous exclusion criteria. Tests of the simple effects found that, contrary to the hypotheses, the interaction was driven by the rude, $F(1, 492) = 5.59, p < .05; \eta^2_p = .01$, rather than the racial, $F(1, 492) = 1.25, p = .25; \eta^2_p = .003$, confrontation condition. Participants in the high (v. low) affiliative motivation condition felt more negative self-directed emotions after receiving a rude ($M = 1.72, SD = 1.00$) compared to racial confrontation ($M = 1.40, SD = 0.85$).

**NegOther.** There was no significant main effect of affiliative motivation or confrontation type on negative other-directed emotions, but there was a significant interaction that did not come out in the more comprehensive sample. The effect mirrored that for negative self-directed emotions, such that the interaction was driven by the rude, $F(1, 492) = 7.55, p < .01; \eta^2_p = .02$, rather than the racial, $F(1, 492) = 1.43, p = .23; \eta^2_p = .003$, confrontation condition. Participants in the high (v. low) affiliative motivation condition felt more negative other-directed emotions after receiving a rude ($M = 2.71, SD = 1.54$) compared to racial confrontation ($M = 2.15, SD = 1.60$).

**Backlash.** With backlash entered as the dependent variable, the two-way ANOVA was like the main analysis, indicating no significant main effect of affiliative motivation, but a significant main effect of confrontation type (see Descriptive Statistics section for more details). The interaction of affiliative motivation and confrontation type on backlash was not significant.
Attitudes Toward Black People. The ANOVA indicated no main effect of affiliative motivation on attitudes toward Black people. Differing from the main analysis, the main effect of confrontation on attitudes was trending significant, such that participants confronted for racism ($M = 5.67, SD = 1.57$) had more positive attitudes toward Black people than those confronted rudely ($M = 5.42, SD = 1.57$). The interaction of affiliative motivation and confrontation type on attitudes was not significant. Moreover, the marginally significant main effect of confrontation type on attitudes toward Mexican, $F(1, 492) = 3.79, p = .05; \eta^2_p = .01$, and Asian, $F(1, 492) = 2.88, p = .09; \eta^2_p = .01$, people followed the same pattern as that for Black people.

Stereotype Endorsement. As in the more comprehensive sample, the two-way ANOVA for stereotype endorsement showed no main effect of affiliative motivation but found a main effect of confrontation type. Participants who received the racial confrontation were more likely to endorse positive stereotypes of Black people ($M = 3.97, SD = 0.88$) than those who received the rude confrontation ($M = 3.79, SD = 0.87$). There was no significant interaction between affiliative motivation and confrontation type.

Intentions to Control Future Bias. Adding to the main analysis, results from the two-way ANOVA with intentions to control future bias as the dependent variable indicated a marginally significant main effect of affiliative motivation. Participants who experienced high affiliative motivation were more likely to say they would monitor their future behaviors for bias ($M = 2.97, SD = 1.76$) than those who experienced low affiliative motivation ($M = 2.66, SD = 1.68$). The main effect for confrontation type was significant, in a way that was again contrary to prior research and the current hypotheses, such that participants who received the racial confrontation were less likely to monitor their future behaviors for bias ($M = 2.64, SD = 1.76$)
than those who received a rude confrontation \((M = 2.98, SD = 1.68)\). The interaction of affiliative motivation and confrontation type on intentions to control future bias was not significant.

Table 22. Supplemental ANOVA results detailing the effects of affiliative motivation and confrontation type on all variables of interest.

<table>
<thead>
<tr>
<th>DV: Trust</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>1.13</td>
<td>1.00</td>
<td>.32</td>
<td>.002</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>0.75</td>
<td>0.67</td>
<td>.41</td>
<td>.001</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.20</td>
<td>0.18</td>
<td>.67</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: NegSelf</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.70</td>
<td>0.74</td>
<td>.39</td>
<td>.002</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>0.22</td>
<td>0.23</td>
<td>.63</td>
<td>.001</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>5.96</td>
<td>6.32</td>
<td>.01*</td>
<td>.013</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: NegOther</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>2.79</td>
<td>1.10</td>
<td>.30</td>
<td>.002</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>6.89</td>
<td>2.70</td>
<td>.10</td>
<td>.005</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>19.43</td>
<td>7.63</td>
<td>.006**</td>
<td>.015</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Backlash</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>1.87</td>
<td>2.73</td>
<td>.10</td>
<td>.006</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>8.11</td>
<td>4.21</td>
<td>.04*</td>
<td>.008</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.03</td>
<td>0.08</td>
<td>.79</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Attitudes Toward Black People</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>1.54</td>
<td>0.63</td>
<td>.43</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>8.11</td>
<td>3.35</td>
<td>.07*</td>
<td>.007</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.03</td>
<td>0.01</td>
<td>.92</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DV: Stereotype Endorsement</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliative Motivation (A)</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>.001</td>
</tr>
<tr>
<td>Confrontation Type (B)</td>
<td>1</td>
<td>4.10</td>
<td>5.40</td>
<td>.02*</td>
<td>.011</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.47</td>
<td>0.62</td>
<td>.43</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Moderated Mediation**

To test hypotheses 1b-4, I separated cases by confrontation type (racial v. rude) and used the PROCESS macro for SPSS which estimates a process model for moderated mediation (Model 8; Preacher et al., 2007) with bootstrapping \((n = 5000)\). Affiliative motivation \((\text{low} = 0 \text{ v. high} = 1)\) served as the independent variable, NegSelf and NegOther were mediators, the combined trust scale served as the moderator, and backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias were tested separately as the dependent variables. The indices for moderated mediation for all models regardless dependent variables were not significant (see Table 23). The following results for each hypothesis are taken from pieces of the overall models.

Table 23. PROCESS Model 8 supplemental results depicting the index of moderated mediation for the full model with affiliative motivation serving as the independent variable, negative self- and other-directed emotions acting as mediators, and backlash and prejudice reduction outcomes as dependent variables.

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>NegSelf</th>
<th>NegOther</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backlash</td>
<td>0.001 (0.01)</td>
<td>[0.02, 0.03]</td>
</tr>
<tr>
<td>Attitudes Toward Black People</td>
<td>-0.004 (0.04)</td>
<td>[-0.09, 0.06]</td>
</tr>
<tr>
<td>Stereotype Endorsement</td>
<td>-0.003 (0.03)</td>
<td>[-0.06, 0.04]</td>
</tr>
<tr>
<td>Intentions to Control Future Bias</td>
<td>0.01 (0.08)</td>
<td>[0.15, 0.18]</td>
</tr>
<tr>
<td><strong>Rude Confrontation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backlash</td>
<td>0.01 (0.01)</td>
<td>[-0.01, 0.03]</td>
</tr>
<tr>
<td>Attitudes Toward Black People</td>
<td>-0.01 (0.02)</td>
<td>[-0.06, 0.03]</td>
</tr>
</tbody>
</table>

**Note:** MS = Mean squares, Effect size = \(\eta^2_p\), \(*p < .05\), **\(p < .01\), \(+p\) is trending significance.
### Hypothesis 1b
I examined the highest order unconditional interactions of affiliative motivation and trust on negative self- and other-directed emotions to see whether trust moderated the strength of the association between affiliative motivation with the confronter and emotional reactions after a racial (v. rude) confrontation. Similar to the main analysis, results did not support hypothesis 1b, finding that trust did not moderate the association between affiliative motivation and negative self-, $F(1, 235) = 0.02, p = .89, R^2 = 0.001$, or other-directed emotions, $F(1, 235) = 1.45, p = .23, R^2 = 0.01$, in the racial confrontation. Likewise, the same results occurred for negative self-, $F(1, 254) = 1.51, p = .22, R^2 = 0.01$, and other-directed emotions, $F(1, 254) = 0.16, p = .69, R^2 = 0.001$, in the rude confrontation condition.

### Hypothesis 2
I sought to replicate past research and test whether negative emotionality mediated the relationship between confrontation content and the outcomes of backlash and prejudice reduction. I predicted that for participants confronted for racism (v. rudely), as negative self-directed emotions increase, backlash and prejudice will decrease and as negative other-directed emotions increase, backlash and prejudice will increase. I examined the beta coefficients and 95% confidence intervals of the model to draw conclusions about significance. The following are results for the effect of negative self-directed emotions on each dependent variable, as negative-other directed emotions did not significantly predict backlash, attitudes toward Black people, stereotype endorsement, or intentions to control future bias in this model.

**Backlash.** These results also supported part of hypothesis 2, as the model indicated that as negative self-directed emotions increased, backlash decreased. The effect in this sample was

<table>
<thead>
<tr>
<th>Stereotype Endorsement</th>
<th>-0.01 (0.01)</th>
<th>[-0.05, 0.01]</th>
<th>0.001 (0.01)</th>
<th>[-0.01, 0.03]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intentions to Control Future Bias</td>
<td>0.05 (0.04)</td>
<td>[-0.03, 0.14]</td>
<td>-0.01(0.02)</td>
<td>[-0.06, 0.04]</td>
</tr>
</tbody>
</table>
also significant for both the racial ($b = 0.10$, 95% CI [0.03 – 0.16]) and rude ($b = 0.07$, 95% CI [0.001 – 0.15]) confrontation conditions. Unlike the main analysis, there was no effect of negative other-directed emotions on backlash.

**Attitudes Toward Black People.** Mirroring the results of the main analysis, as negative self-directed emotions increased, positive attitudes toward Black people significantly decreased ($b = -0.28$, 95% CI [-0.49 – -0.07]) in the racial (v. rude) confrontation condition, suggesting more prejudice. Again, a similar pattern emerged for attitudes toward Mexican ($b = -0.21$, 95% CI [-0.41 – -0.02]) and Indigenous ($b = -0.19$, 95% CI [-0.40 – -0.01]) people, such that as negative self-directed emotions increased, positive attitudes toward these group decreased after a racial (v. rude) confrontation.

**Stereotype Endorsement.** Replicating the main analysis, as negative self-directed emotions increased, positive stereotype endorsement decreased ($b = -0.20$, 95% CI [-0.32 – -0.09]) in the racial (v. rude) confrontation condition. Again, this effect suggests an increase in prejudice rather than the predicted reduction.

**Intentions to Control Future Bias.** Results were similar to the main analysis, such that as negative self-directed emotions increased, intentions to control future bias increased. Again, the effect was significant for both the racial ($b = 0.73$, 95% CI [0.53 – 0.93]) and rude ($b = 0.34$, 95% CI [0.11 – 0.58]) confrontation conditions. However, there was no effect of negative other-directed emotions on intentions to control future bias.

**Hypothesis 3.** The model also tested whether negative emotionality mediated the association between affiliative motivation with the confronter and backlash and prejudice reduction outcomes following a racial (v. rude) confrontation. I examined the model’s indirect effects of negative self- and other-directed emotions on the link between affiliative motivation
and backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias and their 95% confidence intervals to draw conclusions regarding significance.

The results of this sample, using more rigorous exclusion criteria, were identical to those of the main analysis. The association between affiliative motivation and negative self-directed emotions was not significant for either the backlash or prejudice reduction outcomes for either the racial ($b = -0.24$, 95% CI [-1.03 – 0.55]) or rude ($b = -0.11$, 95% CI [-0.83 – 0.61]) confrontation conditions. Likewise, the same non-significant pattern occurred for affiliative motivation predicting negative other-directed emotions for the racial ($b = 0.48$, 95% CI [-0.68 – 1.64]) and rude ($b = 0.82$, 95% CI [-0.26 – 1.90]) confrontations. Hypothesis 3 was not supported as there was no viable link between affiliative motivation and negative emotionality, causing the mediation chain to break down for all dependent variables of interest.

**Hypothesis 4.** The model tested the hypothesis that trust moderated the strength of the association between affiliative motivation with the confronter and backlash and prejudice reduction outcomes after a racial (v. rude) confrontation. To determine significance, I examined the highest order unconditional interactions of affiliative motivation and trust on backlash, attitudes toward Black people, stereotype endorsement, and intentions to control future bias.

Similar to the main analysis, the results indicated that trust did not moderate the association between affiliative motivation and backlash for either the racial, $F (1, 233) = 0.07, p = .79, R^2 = 0.001$, or rude, $F (1, 252) = 1.07, p = .30, R^2 = 0.001$, confrontation conditions. Unlike the main analysis, the results remained non-significant regardless of the type of trust that was entered as the moderator (i.e., outgroup or workplace).

In terms of prejudice reduction, trust again did not moderate the association between affiliative motivation and attitudes toward black people (racial: $F (1, 233) = 0.19, p = .67, R^2 =$...
With the more rigorous exclusion criteria, there was only a marginally significant result for intentions to control future bias. The pattern of results showed such that trust moderated the association between affiliative motivation and intentions to control future bias at low levels of trust ($b = 0.67, 95\% \text{ CI} [0.09 – 1.26]$) for the racial, $F(1, 233) = 3.61, p = .06, R^2 = 0.01$, but not rude, $F(1, 252) = 0.45, p = .50, R^2 = 0.001$, confrontation. In other words, this finding suggests that for participants who felt low affiliative motivation, as trust increased intentions to control future bias increased. The impact of trust on the association between affiliative motivation and intentions to control future bias were not significant for those in the high affiliative motivation condition (see Table 24 and Figure 9).

Table 24. PROCESS Model 8 supplemental results depicting the interaction of affiliative motivation and trust on intentions to control future bias after a racial confrontation.

<table>
<thead>
<tr>
<th>Racial Confrontation</th>
<th>Intentions to Control Future Bias</th>
<th>[LLCI, ULCI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>[LLCI, ULCI]</td>
</tr>
<tr>
<td><strong>Affiliative Motivation (AM)</strong></td>
<td>1.30 (0.59)**</td>
<td>[0.53, 0.93]</td>
</tr>
<tr>
<td>NegSelf</td>
<td>0.73 (0.10)**</td>
<td>[0.53, 0.93]</td>
</tr>
<tr>
<td>NegOther</td>
<td>0.08 (0.07)</td>
<td>[-0.06, 0.23]</td>
</tr>
<tr>
<td>Trust</td>
<td>0.55 (0.13)**</td>
<td>[0.29, 0.81]</td>
</tr>
<tr>
<td><strong>AM x Trust</strong></td>
<td>-0.34 (0.18)$^+$</td>
<td>[-0.70, 0.01]</td>
</tr>
</tbody>
</table>

**Conditional effects of AM x Trust on Intentions to Control Future Bias at three levels of trust.** Analyses use continuous variables; effects below are one standard deviation +/- the mean.

<table>
<thead>
<tr>
<th>Trust</th>
<th>B (SE)</th>
<th>[LLCI, ULCI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ($M = 1.82$)</td>
<td>0.67 (0.30)$^*$</td>
<td>[0.09, 1.26]</td>
</tr>
<tr>
<td>Medium ($M = 3.08$)</td>
<td>0.24 (0.19)</td>
<td>[-0.14, 0.62]</td>
</tr>
<tr>
<td>High ($M = 4.17$)</td>
<td>0.14 (0.28)</td>
<td>[-0.69, 0.41]</td>
</tr>
</tbody>
</table>

*Note:* $^*$ $p < .05$, $^{**}p < .01$, $^{+}p = 0.06$
Study 2: Supplemental Results Summary

After excluding the additional 372 participants for not passing the manipulation check, the supplemental analyses revealed some notable differences compared to the main analyses. Participants in the high affiliative motivation condition reported being more connected to their interaction partner and alert compared to those in the low affiliative motivation condition. Affiliative motivation was correlated with intentions to control future bias, and those in the high (v. low) affiliative motivation condition reported more intentions to monitor themselves for bias.

However, there was no longer a correlation between confrontation type and negative other-directed emotions. Likewise, the main effect of confrontation type on negative other-directed emotions was no longer significant in these supplemental analyses. Instead, an interaction between affiliative motivation and confrontation type on negative self- and other-
directed emotions appeared. The pattern of the result for these interactions was not in line with hypotheses.

Finally, trust (regardless of measure) no longer moderated the association between affiliative motivation and negative other-directed emotions or affiliative motivation and backlash. The previously significant moderating effect of trust on the association between affiliative motivation and intentions to control future bias was only marginally significant.

While these analyses revealed an effect of affiliative motivation on feeling connected with one’s interaction partner in the predicted direction, it came with the risk of, once again, violating random assignment and rendering conclusions mildly interpretable. Additionally, with less power to detect effects, integral results from the main analyses were rendered insignificant.


Blieszner, R. (2014). The worth of friendship: Can friends keep us happy and healthy?


Cable, N., Bartley, M., Chandola, T., & Sacker, A. (2013). Friends are equally important to men and women, but family matters more for men's well-being. Journal of Epidemiology Community Health, 67(2), 166-171.


VITA

Dr. Jamie Patrianakos was born and raised in the southwest suburbs of Chicago, Illinois. Before attending Loyola University Chicago, she earned a Bachelor of Arts in Psychology, with Highest Distinction, at The Pennsylvania State University in 2012. She received her Master of Arts in Social Psychology from Loyola University Chicago in 2018.

While at Loyola, Dr. Patrianakos served on the Department of Psychology’s Committee on Diversity Affairs and Graduate Diversity Orientation Committee. At the university level, Dr. Patrianakos was the co-chair of the organization, Enhancing Diversity in Graduate Education, and a volunteer member of the Black Lives Matter conference and the search committee for Loyola’s Vice President of Institutional Diversity, Equity, and Inclusion.

Dr. Patrianakos has won several awards from the Psychology Department, including the Victor J. Heckler Fellowship and the Social Psychology Graduate Student Research and Professional Development Scholarship. She has also been awarded the Interdisciplinary Graduate Research Symposium Diversity Award from the Graduate School.

Additionally, Dr. Patrianakos has served as the senior compliance assistant for Loyola’s Office of Research Services. She has taught, and guest lectured courses on the social bases of behavior with a specific focus on the psychology of prejudice. Dr. Patrianakos is currently a Research Health Science Specialist at Edward Hines, Jr. VA Hospital.