The (In)egalitarian Self: On the Motivated Rejection of Implicit Racial Bias

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Joseph A. Vitriol

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Eugene Borgida, Ph.D., Advisor

Mark Snyder, Ph.D., Advisor

October 2016

Acknowledgements

I am grateful for the invaluable contributions and ideas from the many individuals who have supported me throughout the process of my graduate training and doctoral dissertation. Foremost among them are my advisors, Eugene Borgida and Mark Snyder, whose sound advise, thoughtful and challenging questions, and high expectations and standards for excellence have served as a benchmark to which I've aspired throughout my graduate school career. They, along with the other members of my committee, Christopher Federico and Howard Lavine, deserve special recognition for their tremendous support, patience, and superb insight and feedback at every step of this endeavor. Each have served as inspiring mentors, advisors, and teachers to many emerging psychologists, and I too owe a sincere debt of gratitude for having benefited from their guidance. I would also like to thank the many faculty members in the Psychology and Political Science Departments at the University of Minnesota for sharing with me their unique perspectives and invaluable resources that continue to shape my thinking. Thank you for challenging me to grow into a better scientist and a better person. Finally, I would like to thank my colleagues, friends, and family for their ongoing support, love, and much needed humor.

This dissertation would not be possible without the tireless work of my dedicated team of research assistants and many other young scholars that I've been privileged to work with over the years. I am humbled by their generosity and proud of their commitment to the pursuit of scientific knowledge on behalf of the public good. Funding from the Center for the Study of Political Psychology and the Department of Psychology Research Award graciously supported my dissertation work.

The mistakes that remain in this document are my own.

Dedication

I dedicate this dissertation to all of my family, loved ones, and friends, but especially to my parents, Carolyn and Jonathan Vitriol. I am hopeful that this dissertation is the first step of many in repaying the love, support, and patience you have always shown me throughout my life. Thank you for teaching me the courage to pursue my passions and my dreams.

Abstract

White Americans widely endorse egalitarian values and strongly oppose hostile forms of racial prejudice, yet significant racial disparities persist in many important life domains. Unconscious, implicit forms of bias at the individual-level have been offered as one explanation for aggregate racial disparities. Research has identified strategies to increase prejudice-regulation and reduce implicit bias in applied contexts, but has neglected to consider various psychological obstacles to the successful implementation of these interventions. Across three studies in which I experimentally manipulated exposure to scientific information and personalized feedback about implicit bias, I examined one such obstacle: that evidence of implicit racial bias threatens individuals' egalitarian self-concepts, and activates motivated reasoning processes that bolster the denial of implicit bias feedback and its influence on behavior. I also test several strategies to decrease defensive responding and attenuate motivated reasoning in this context. Results indicate that exposing White Americans to credible information on the science of implicit bias *can* increase awareness, but also risks backfire effects in the form of negative attitudes towards social science and increased racial stereotyping. Additionally, personalized implicit bias feedback reliably induced negative attitudes towards an instrument designed to measure implicit bias (i.e., IAT) and negative self-reported affect, but had no reliable influence on awareness. However, negative affect mediated the relationship between feedback and self-perceived bias, suggesting that personalized feedback can have indirect effects on awareness. Importantly, I also obtain robust evidence for the success of a simple pre-feedback intervention informing participants that implicit bias is common and fundamental to human cognition, but nonetheless malleable and subject to control. This collective bias intervention, when paired with personalized feedback, reliably increased self-perceived bias, belief in prejudice and discrimination, and more acceptance of and favorable attitudes towards the IAT. Across all studies, I find that the motivated rejection of implicit bias is consequential for prejudice-regulation, stereotyping, and public policy attitudes, and mediates the relationship between a broad range of individual differences (i.e., sociopolitical orientations, explicit racial attitudes, and egalitarian motivations) and these outcomes. Finally, defensive responding to implicit bias evidence and feedback was cognitively depleting, as indexed by the Stroop test, although impaired performance on this task was reduced among participants in the pre-feedback intervention. I conclude with a discussion of the implications of this evidence for anti-bias interventions, models of prejudice-regulation, impression management strategies in the context of intergroup relations, and the study of racial attitudes and prejudice. I also consider the application of my results to political and legal contexts, and identify future directions for additional research.

Table of Contents

` Page

Acknowledgements i

Dedication iii

Abstract iv

Table of Contents v

List of Tables ix

List of Figures x

Introduction 1

Thesis Overview 7

Evidence for Racial Discrimination and Inequality 11

Old Racism in a New World 13

Implicit Racial Bias and Explicit Egalitarianism 19

Unprejudiced Self-Image Maintenance 21

Motivation to Control and Regulate Prejudiced Responding 25

Current Research: Overview of Theory, Hypotheses, Methods 31

Sociopolitical Orientations and Perceptions of Racial Discrimination 31

Intrinsic vs. Extrinsic Egalitarianism and Prejudice-Related Discrepancies 32

Downstream Consequences of the Motivated Rejection of Implicit Bias 37

Interventions to Attenuate Motivated Rejection of Implicit Bias 38

Overview of Study 1, 2, and 3 40

General Overview of Analysis Across Studies 41

Pilot 1a: General Belief in the Existence of Implicit Racial Bias 42

Goals and Hypotheses 42
Method 43
Results 48
Discussion 49
Pilot 1b: General Belief in the Existence of Implicit Racial Bias 50
Goals and Hypotheses 50
Method 51
Results 52
Discussion 54
Study 1: General Belief in the Existence of Implicit Racial Bias 56
Goals and Hypotheses 56
Method 58
Results 64
Discussion 71
Pilot 2: Allegations of Implicit Bias 78
Goals and Hypotheses 78
Method 79
Results 87
Discussion 94
Study 2: Allegations of Implicit Bias 95
Goals and Hypotheses 95
Method 97

Results 99

Discussion 114

Pilot 3: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback 119

Goals and Hypotheses 119

Method 119

Results 123

Discussion 125

Study 3a: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback 126

Goals and Hypotheses 126

Method 127

Results 128

Discussion 133

Study 3b: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback 135

Goals and Hypotheses 135

Method 136

Results 136

Discussion 143

General Discussion 147

How do White Americans respond to information and feedback about implicit racial bias? 148

Does the motivated rejection of implicit racial bias challenge existing models of prejudice-regulation? 153

Direct Implications for Political and Legal Contexts 163

Future Directions 166

Conclusion 170

References 171

Tables 196

Figures 206

Appendix 243

List of Hypotheses 243

Measures Index 244

Manipulation Checks 247

Pre-Manipulation Measures 247

Post-Manipulation Measures 254

Pilot/ Study 1 Post-Manipulation Questions 268

Study 3 Interventions 269

Open-Ended Coding Scheme for Pilot 2 274

List of Tables

Page

Table 1.1. Means, SD, and alpha for measures used in Pilot 1b 195

Table 1.2. Correlations between all continuous variables used in Study 1 196

Table 2.1. Means, SD, and alpha for measures used in Pilot 2 197

Table 2.2. Correlations between all continuous variables used in Pilot 2 198

Table 2.3. Inter-rater agreement for open-ended content coding of self-reported belief in the use of deception, prior to consensus, for Pilot 2 199

Table 2.4. Correlations between all continuous variables used in Study 2 200

Table 3.1. Means, SD, and alpha for measures used in Pilot 3 201

Table 3.2. Correlations between all continuous variables used in Pilot 3 202

Table 3.3. Means and SD for measures used in Study 3a 203

Table 3.4. Means and SD for measures used in Study 3b 204

List of Figures

Page

Figure 1.1. Main effect of experimental condition on motivated reasoning for Pilot 1a 205

Figure 1.2. Main effect of experimental condition on motivated reasoning for Pilot 1b 206

Figure 1.3. Coefficients plot for effect of T1 independent variables on motivated reasoning in Study 1 207

Figure 1.4. Coefficients plot for effect of T1 independent variables on prejudice-regulation in Study 1 208

Figure 1.5. Coefficients plot for effect of T1 independent variables on stereotyping and public policy attitudes in Study 1 209

Figure 1.6. Experimental condition x social dominance orientation, system justification, and explicit black attitudes on stereotype endorsement. 210

Figure 2.1. Main effect of experimental condition on affect for Pilot 2 211

Figure 2.2. Main effect of experimental condition on motivated reasoning for Pilot 2 212

Figure 2.3. Coefficients plot for effect of affect on motivated reasoning in Pilot 2 213

Figure 2.4. Main effect of experimental condition on affect for Study 2 214

Figure 2.5. SDO x experimental condition on affect for Study 2 215

Figure 2.6. Coefficients plot for effect of affect on motivated reasoning in Study 2 216

Figure 2.7. Main effect of experimental condition on motivated reasoning for Study 2 217

Figure 2.8. Coefficients plot for effect of individual differences on motivated reasoning in Study 2 218

Figure 2.9. SDO x experimental condition on belief in prejudice and discrimination for Study 2 219

Figure 2.10. Egalitarianism x experimental condition on perception of IAT for Study 2 220

Figure 2.11. Egalitarianism x experimental condition on belief in prejudice and discrimination for Study 2 221

Figure 2.12. The relationship between attitudes towards blacks and attitudes towards anti-bias interventions, separated by experimental condition 222

Figure 2.13. Coefficients plot for effect of individual differences on prejudice-regulation in Study 2 223

Figure 2.14. Coefficients plot for effect of individual differences on stereotyping and policy attitudes in Study 2 224

Figure 2.15. The relationship between system justification and attitudes towards anti-bias interventions, separated by experimental condition 225

Figure 3.1. Coefficients plot for effect of affect on motivated reasoning in Pilot 3 226

Figure 3.2. Main effect of experimental condition on affect for Study 3a 227

Figure 3.4. Attitudes towards blacks x experimental condition on affect for 228

Study 3a

Figure 3.5. Intrinsic-extrinsic egalitarianism x experimental condition on negative affect for Study 3a 229

Figure 3.6. Intrinsic-extrinsic egalitarianism x experimental condition on self- and other-directed negative affect for Study 3a 230

Figure 3.7. SDO x experimental condition on self- and other-directed negative affect for Study 3a 231

Figure 3.8. System Justification x intervention condition on self-perceived bias, stereotyping, and policy attitudes for Study 3a 232

Figure 3.9. Main effect of experimental condition on affect for Study 3b 233

Figure 3.10. Main effect of experimental condition on Motivated Reasoning for Study 3b 234

Figure 3.11. Main effect of experimental condition on Motivated Reasoning for Study 3b 235

Figure 3.12. Egalitarianism x experimental condition on belief in prejudice for Study 3b 236

Figure 3.13. SDO x experimental condition on belief in prejudice for Study 3b 237

Figure 3.14. SDO x experimental condition on perception of IAT for Study 3b 238

Figure A. Conceptual Model 239

The (In)egalitarian Self: On the Motivated Rejection of Implicit Racial Bias

Racial prejudice has a long and nefarious history in America. To be sure, progress under the banner of tolerance and egalitarianism has undoubtedly been made, and has led to meaningful changes in the ways in which people interact with members of traditionally disadvantaged groups. The days of lynching, segregation, or other unlawful and egregious public displays of bigotry are surely less common, and in its wake proactive efforts to embrace and instill values of racial equality have become more prevalent. While the expression and institutionalization of racial animosity is rarely countenanced among the more thoughtful members of society, its social and psychological legacy has nonetheless endured.

In contemporary society, Americans widely endorse egalitarian values and strongly oppose racially biased policy and legislation. Indeed, most people generally reject hostile forms of prejudice and are motivated to downplay racial bias in their own judgment and behavior (Dunton & Fazio, 1997), preferring to view themselves as egalitarian and society as fair (Bobo, 2001; Sears, Henry & Kosterman, 2000). Although the rise of egalitarian values and norms have increased the social and moral costs of appearing or being prejudiced, it has done little to dismantle or forego a race-based hierarchy that relegates racial minorities to inferior status within society (Bobo, Klugel, & Smith, 1996). Significant and consequential racial disparities continue to persist across a broad range of contexts (Blank, 2001; Fisher & Borgida, 2012).

Considerable evidence in the social sciences has established the existence of racially biased *implicit* attitudes, which have been offered as an explanatory account for racial disparities in many important life domains (Greenwald & Krieger, 2006). Implicit

attitudes are associations that, unlike explicit attitudes, may not be accessible through deliberate introspection, but operate quickly on judgment and behavior without conscious awareness, effort, or control (Gawronski, Hofmann, & Wilbur, 2006). Thus, implicit attitudes can bias judgment and behavior in racially inequitable ways that may be inconsistent with individuals' explicit or consciously accessible attitudes and values (Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

Despite Americans' general commitment to antidiscrimination principles and values (O'Brien, Crandall, Horstman-Reser, Warner, Alsbrooks, & Blodorn, 2010), common understanding of contemporary forms of discrimination anchor on explicitly hostile, "traditional" forms of prejudice, which are prohibited by law (Kang & Banaji, 2006). Accordingly, many organizations have attempted to implement strategies to reduce explicit forms of racial discrimination (Kalev, Dobbin, & Kelly, 2006). However, the existence of unconscious, implicit racial bias fundamentally challenges the effectiveness of these more conventional efforts that focus exclusively on conscious attitudes to reduce discrimination (Krieger, 2008). Individuals, organizations, and institutions committed to antidiscrimination principles would be remiss to ignore the empirical evidence on implicit racial bias and its relation to discriminatory outcomes (Jost, Rudman, Blair, Carney, Dasgupta et al., 2009).

For example, a claim of disparate impact in public policy and hiring practices may be legally actionable, even in the absence of explicit discriminatory intent (Bartlett, 2009; *Texas Department of Housing and Community Affairs vs. Inclusive Communities Project*, 2015). More generally, racially inequitable treatment by the criminal justice system can undermine the public's trust, perceived legitimacy, and compliance with law

enforcement, especially among racial minorities (Peffley & Hurwitz, 2010). Similarly, the strategic use of subtle racial cues in political communication can be successful in mobilizing parts of the electorate characterized by unconscious racial bias (Mendelberg, 2001), which may contribute to the principled opposition to policies designed to improve the lives of racial minorities (e.g., social welfare; Gilens, 2000). Efforts to attenuate implicit racial bias (in addition to explicit or overt bias) are therefore a necessary reform to improve the public's trust in and acceptance of legal and political processes, institutions, and outcomes.

The existence of racial disparities and its relation to implicit bias have prompted an increasing number of promising empirical investigations into strategies that can reduce such biases and their consequences (Devine, Monteith, Zuwerink, & Elliot, 2012; Smedley, Stuth, & Nelson, 2003; Lai, Marini, Lehr, Cerruti, Shun, Joy-Gaba et al., 2015). Nevertheless, this work has not considered the extent to which additional psychological obstacles may undermine the effective implementation of these interventions. For example, many White Americans, even those who identify as egalitarian, may be threatened by the idea that subtle, unconscious prejudice can influence their behavior, without their knowledge or ability to control it, and in ways that contributes to racial disparities that are inconsistent with their explicit values and beliefs. The mere possibility that one is implicitly racially biased may be perceived as a failure or unwillingness to comply with important social norms- an indictment of egalitarian motives and the integrity of one's personal character. Even White Americans' who are receptive to the idea that racial discrimination and inequality exists in society may struggle to accept the possibility that they too harbor racial bias.

Consistent with this idea, a substantial body of research indicates that the sociopolitical orientations of many White Americans render them incapable or unwilling to recognize patterns of racial inequality and discrimination (Adams, Tormala, & O'Brien, 2006). Even self-identified egalitarians are fundamentally threatened by and avoidant of situations, interactions, and information that may challenge their egalitarian self-image (Dovidio & Gaertner, 1986; Frantz, Cuddy, Burnett, Ray, & Hart, 2004; Gaertner & Dovidio, 1986, 2000). Similarly, people engage in a range of cognitive strategies to protect and maintain the unprejudiced self-image, including self-serving social comparisons, symbolic endorsement of "color-blind" ideologies, and narrow definitions of what constitutes racial prejudice (O'Brien et al., 2010; Norton, Vandello, Biga, & Darley, 2008; Sommers & Norton, 2006). These strategies for persuading the self and others of one's egalitarianism allows for the rationalization of socially unacceptable manifestations of racial prejudice, including its implicit forms. Unfortunately, increasing self-awareness of racial bias is both an important starting point for the effective regulation of prejudice expression (Monteith & Mark, 2009), and a common feature of anti-prejudice interventions (e.g., Blair, 2002; Banaji, 2001; Paluck & Green, 2009). Consequently, defensiveness and resistance, not increased awareness, may be a common reaction to evidence of racial discrimination and interventions designed to reduce implicit racial bias.

More importantly, this defensiveness may activate motivated reasoning processes (e.g., Kunda, 1990) that undermine acceptance of evidence for and allegations of implicit racial bias. The motivated denial of implicit racial bias may bolster individuals' unwillingness to comply with or engage in anti-prejudice interventions, or adjust for the

effect of implicit racial bias on their judgment and behavior. In this way, strategies to reduce implicit racial bias via education and self-awareness may risk backfire effects. The motivated rejection of allegations and evidence of implicit racial bias and its consequences for behavior therefore constitute a potential obstacle to effective prejudice-regulation and the successful implementation of interventions intended to reduce implicit bias. Yet understanding this phenomenon and identifying strategies to reduce resistance has been conceptually and empirically neglected.

Specifically, prior work has neglected to study the psychological conditions under which people are receptive to evidence and allegations of implicit racial bias (see Howell et al., 2015 for an exception). Only some research has examined aversion and avoidance among White Americans to information concerning one's implicit racial bias. For example, Howell et al. (2013) find that people are avoidant of feedback that might suggest they harbor unconscious racial biases. When directly confronted with accurate feedback suggesting that one's implicit attitudes are more racially biased than explicitly held beliefs, people are distressed and respond defensively (Howell et al., 2015). Several limitations of this work, however, minimize its value for the questions explored in the current research. For instance, Howell and colleagues (2015) primarily adopted a correlational design, which precludes strong causal inferences regarding the impact of implicit bias feedback on motivated reasoning processes, and its implications for stereotype endorsement, prejudice-regulation, and other outcomes of interest. Furthermore, the role of sociopolitical orientations, as described above, and the motivation to control prejudice in patterning White Americans' response to implicit bias feedback remains unexamined.

Prior research also has not explored the consequences of the motivated rejection of implicit bias for the willingness to reduce its impact on one's behavior and other downstream effects. In my thesis research, I investigate the hypothesis that evidence of one's own implicit racial bias can threaten individuals' egalitarian and moral sense of self. More specifically, it is expected that the possibility of one's own implicit racial bias will be experienced as particularly distressing among White Americans, even those who strongly endorse egalitarian values or view racial intolerance as a moral imperative (Dovidio & Gaertner, 1986; Eberhardt & Fiske, 1998; Greenwald & Banaji, 1995). This threat to the self, however, may have different consequences and implications for White Americans, depending on the extent to which egalitarian social norms and personal standards for behavior have been internalized into their self-concepts.

For instance, defensiveness and motivating reasoning is particularly likely to characterize the reaction of White Americans' who are motivated to *project* an egalitarian self-image for extrinsic reasons (e.g., to avoid the attribution of racist intent) or primarily to comply with prevailing egalitarian social norms. Highly prejudiced, extrinsically motivated egalitarians commonly respond with resentment and denial to information suggesting that they have acted in ways inconsistent with egalitarian social norms (Monteith, 1993; Strauman & Higgins, 1987). In contrast, intrinsically motivated White Americans' who have internalized egalitarian standards for their own personal conduct commonly response to prejudice-related discrepancies with guilt and compunction (Devine et al., 1991; Monteith & Voils, 1998; Zuwerink et al., 1996). This self-directed negative affect has been found to activate prejudice-regulation (Devine et al., 2002), and could therefore increase acceptance of evidence and allegations of implicit racial bias.

However, it remains unclear and unexamined if implicit bias feedback and evidence operates in a similar fashion as other kinds of prejudice-related discrepancies.

The motivated rejection of evidence in the face of allegations of implicit racial bias and its relationship to aggregate-level racial disparities may have several important downstream consequences not considered by prior research. For example, it may increase personal (vs. situational) attributions for the causes of aggregate-level racial disparities (e.g., Peffley & Hurwitz, 2010), which, in turn, could increase support for more punitive attitudes towards the criminal justice system and opposition to policy remediation to address racial inequality (e.g., affirmative action and social welfare). Furthermore, the effects of these motivated reasoning processes are also expected to undermine individuals' perceived value and importance of organizational interventions designed to reduce implicit racial bias**,** and their general motivation to control prejudiced responding and its impact on behavior (Dunton & Fazio, 1997).

For my dissertation, I test and find evidence largely consistent with these hypotheses across three studies in which I manipulate *general* beliefs about the existence of implicit bias (Study 1) and *specific* beliefs regarding one's own personal implicit bias (Study 2). This approach will allow me to assess the causal effects of information about and allegations of implicit racial bias on motivated reasoning processes, as well as strategies to attenuate this defensiveness and its impact on behavior (Study 3).

Thesis Overview

In the sections that follow, I describe my theoretical framework from which these predictions are derived. First, I review the empirical evidence for aggregate-level racial inequality and discrimination, and the link between racial attitudes, stereotype

endorsement, and opposition to public policies intended to help racial minorities. I then consider the implications of persistent racial inequality and discrimination despite the rise of egalitarian values. I argue that the discrepancy between racially tolerant values and the acceptance of existing racial disparities indicates the emergence of a new form of racial prejudice.

This attitudinal orientation, however, cannot clearly be understood in terms of traditional, more deliberate forms of hostile racial prejudice. Overt, hostile expressions of prejudicial attitudes towards racial minorities have declined substantially in recent decades (Dovidio & Gaertner, 1998). Instead, egalitarian social norms have promoted a more subtle form of resentment towards racial progress, opposition to public policy intended to help racial minorities, and complaisance with systemic inequality through which racial prejudice is justified in socially acceptable terms. This analysis implies that the sociopolitical orientation of many White Americans' render them unable or unwilling to recognize or address aggregate-level patterns of racial disparities. Consequently, many White Americans will be motivated to justify or reject *all* evidence of racial inequality and discrimination, including implicit racial bias.

More importantly, egalitarian social norms have increased the social and moral cost of appearing and being prejudiced, while leaving the psychological legacy of racial hierarchy largely intact. Consistent with this perspective, I review evidence for the existence of automatic, unconscious forms of racial bias, which has been found to impact judgment and behavior at the individual-level in ways that may be socially unacceptable and inconsistent with even sincerely held egalitarian values. Thus, egalitarian social norms may have attenuated the hostile expression of racial prejudice, but unconscious

forms of racial bias nonetheless persist and may directly contribute to aggregate-level disparities. Consequently, many White Americans are threatened by the prospect of being judged as racist for behaviors or beliefs they are unaware of and may not be able to control, which heightens the psychological need to maintain an unprejudiced self-image and the social costs of appearing otherwise. Accordingly, I review evidence that supports the proposition that many White Americans are averse to and avoidant of information that may threaten their egalitarian self-image, or may otherwise engage in a broad range of social and psychological strategies to protect their unprejudiced self-concepts.

The introduction ends with a review of research on the motivation to regulate and control the expression of prejudice. This work suggests that White Americans' who have internalized egalitarian standards for their personal conduct (i.e., intrinsic egalitarians) are both motivated and able to respond to evidence of prejudice-related discrepancies with compunction which in turn increases efforts to regulate their prejudiced thoughts and behaviors. I review literature emphasizing the importance of self-awareness of prejudice related-discrepancies for affect and the motivation to control and regulate prejudice responding, and its implications for the psychological conditions under which individuals are expected to be receptive to feedback suggesting that they harbor implicit racial bias.

After describing in great detail my hypotheses and methods, I then report the results of 4 pilot studies and 4 experiments across 3 studies, involving approximately 2400 White Americans, in which I investigate the psychological factors implicated in the motivated rejection of implicit racial bias evidence and feedback, and its downstream consequences. Study 1 was designed to manipulate the belief in the existence of implicit

bias by providing participants' with credible information that summarizes psychological research and theory on implicit cognition and racial discrimination. Study 2 will experimentally manipulate participants' belief in their own implicit bias using a false-feedback paradigm following completion of the Implicit Association Test (Greenwald, McGhee, & Schwartz 1998). Finally, in Study 3, I will independently manipulate "self-affirmation" and "collective bias" pre-feedback interventions to evaluate the relative success of each in reducing motivated reasoning processes and its effects.

The results suggest that exposing White Americans to credible information on the science of implicit bias *can* increase awareness, but also risks backfire effects in the form of negative attitudes towards social science and increased racial stereotyping. Additionally, personalized implicit bias feedback reliably induced negative attitudes towards the IAT and negative self-reported affect, but had no reliable influence on awareness. However, negative affect mediated the relationship between feedback and self-perceived bias, suggesting that personalized feedback can have indirect effects on awareness. Importantly, I also obtain robust evidence for the success of a simple pre-feedback intervention informing participants that implicit bias is common and fundamental to human cognition, but nonetheless malleable and subject to control. This collective bias intervention, when paired with personalized feedback, reliably increased self-perceived bias, belief in prejudice and discrimination, acceptance of the IAT results, and more favorable attitudes towards the IAT. Across all studies, I find that the motivated rejection of implicit bias is consequential for prejudice-regulation, stereotyping, and public policies attitudes, and mediates the relationship between a broad range of individual differences (i.e., sociopolitical orientations, explicit racial attitudes, and

egalitarian motivations) and these outcomes. Finally, defensive responding to implicit bias feedback was cognitively depleting, as indexed by the Stroop test, although impaired performance on this task was reduced among participants in the pre-feedback intervention. I conclude with a discussion of the implications of this evidence for anti-bias interventions, models of prejudice-regulation, impression management strategies in the context of intergroup relations, and the study of racial attitudes and prejudice. I also consider the application of my results to political and legal contexts, and identify future directions for additional research.

Evidence for Racial Discrimination and Inequality

Evidence for racially disparate treatment and outcomes across most important life domains abounds. Ranging from even the most ordinary interpersonal interactions (e.g., Sue, Capodilupo, Torino, Bucceri, Holder, Nadal, & Esquilin, 2007) to more consequential outcomes in health contexts, employment and hiring decisions, education, criminal justice contexts, housing opportunities, and more, racial minorities are commonly, and systematically, bedeviled by disfavor, maltreatment, and disadvantage (e.g., Alexander, 2010; Brief et al., 2000; Bobo, Kluegel, & Smith, 1996; Huddy & Feldman, 2009; National Research Council, 2004; Oliver & Shapiro, 1995; Peffley & Hurwitz, 2010; Roth, Huffcutt, & Bobko, 2003; Yinger, 1998; Zeigert & Hanges, 2005). While mean level differences in aggregate-outcomes do not directly indicate evidence of discriminatory intent or effects (National Research Council, 2004), research that controls for race-neutral factors that may contribute to observed disparities also supports the conclusion that minorities are commonly treated less fairly and judged based on different standards than their White counterparts.

In particular, research examining discrimination in employment decisions and mortgage lending practices have observed clear patterns of racial disparities, even when statistical controls for job relevant factors, credit history, and financial resources are included (e.g., Bendrick, Jackson, & Reinoso, 1994; Leslie et al., 2008; Oppler, Campbell, Pulakos, & Borman, 1992). Similarly, significant racial differences in the rates of arrest, prosecution, incarceration, or sentencing outcomes can not be explained by differential rates of criminality among racial minorities (e.g., Alexander, 2010; Lauritsen & Sampson, 1998; Macdonald, 2001; Mauer, 1999; Peffley & Hurwitz, 2010; Walker, Spohn & DeLone, 2004;Western, 2006; Yates & Fording, 2005). Furthermore, field experiments and audit studies, in which race-neutral variables are held constant and the race of applicants is manipulated experimentally, provide even stronger evidence of systematic discrimination in many real-world contexts involving actors unaware of their participation in an experiment (e.g., Altonji & Blank, 1999; Ayres, 2001; Bendick & Nunes, 2012; Doleac & Stein, 2013; Pager, 2007; Quillian 2006; Yinger, 1993). For example, Bertrand & Mullainathan (2003) varied the names of applicants for employment in Chicago and Boston to be stereotypically black or white. Although job applicants' resumes were matched to be identical, applicants with stereotypically white names received 50% more inquires from prospective employers than applicants with stereotypically black names.

Perhaps the strongest evidence for the pervasiveness and consequentiality of racial discrimination in contemporary society is research that focuses on the direct link between racial attitudes and support for public policies intended to help racial minorities. For example, negative racial attitudes have been found to undermine support for affirmative

action, social welfare, interracial busing, healthcare reform, less punitive criminal justice policies and minority political candidates (Abramowitz, 1994; Federico & Sidanius, 2002; Gilens, 1995; Huddy & Feldman, 2006, 2009; Kinder & Sanders, 1996; Kluegel, 1990; Sears et al., 1997; Sears et al., 2000; Sidanius, Pratto, & Bobo, 1996; Tesler, 2012, 2013; Tesler & Sears, 2012; Virtanen & Huddy, 1998). These effects have been observed above and beyond the impact of general ideological principles, such as individualism and political conservatism (Kinder & Mendelberg, 2000; Sears & Henry, 2003; Reyna et al., 2005). Consistent with these findings, experimental research provides additional evidence that public policy support varies predictably as a function of the race of the group most likely to benefit, such that whites are less supportive of policies intended to help racial minorities (Bobo & Kluegel, 1993; Huddy & Feldman, 2009; Hurwitz & Peffley, 1997; Sniderman et al., 1996). Opposition to public policy intended to help racial minorities is particularly disingenuous given the robust evidence of aggregate-level discrimination and disadvantages experienced within these communities.

Thus, across diverse methodologies and participant samples, convergent evidence from decades of research clearly supports the existence of racial discrimination in many important life domains that cannot be explained by race-neutral factors. Furthermore, this work also supports a strong link between explicit racial attitudes and discriminatory preferences for public policy. Together, this body of research seriously calls in to question the credibility of race-neutral explanations for gross racial disparities across many important life domains.

Old Racism in a New World

The persistence of discriminatory outcomes for racial minorities and its relation to

explicit racial attitudes can be contrasted with the rise of more egalitarian values and racially tolerant social norms in contemporary times (Bobo, Kluegel, & Smith, 1996; Crandall, Eshleman, & O'Brien, 2002; Katz, Wackenhut, & Hass, 1986; O'Brien et al., 2010; Page & Shapiro, 1992; Pettigrew, 2003; Plant & Devine, 2001). Belief in racial inferiority and overt animosity towards racial minorities are no longer acceptable nor common among the public[1]. Support for state sanctioned segregation and discrimination is indeed a relic of an old social order long in disrepute. Contemporary Americans largely believe that racial prejudice is personally unacceptable (Crandall et al., 2002; Plant & Devine, 1998), and have internalized egalitarian standards that guide their personal behavior (Monteith, 1993).

In this section, I argue that these dramatic changes in public attitudes on race, however, have not necessarily resulted in the principled rejection of racial inequality, nor the outright condemnation of racial stereotypes as an explanation for disadvantaged standing among racial minorities (Bobo, Kluegel, & Smith, 1996; Sommers & Norton, 2011). As already noted, there remains a large disconnect between support for egalitarian racial principles and support for egalitarian social policies designed to improve the living conditions and prospects of racial minorities (Schuman, Steeh, & Bobo, 1985; Sears et al., 1997)**.** Furthermore, White Americans are both threatened by and inclined to over-estimate progress on racial equality. As a result, many White Americans are motivated to believe that improved economic and social standing among racial minorities comes at a cost to their own social and economic standing (Bobo, Kluegel, & Smith, 1996; Dixon et

[1] Although, a sizable minority of White Americans (20-25%) still believe that racial inequality can largely be explained by genetic differences in intelligence or other characteristics (Huddy & Feldman, 2006, 2009).

al., 2007; Eibach & Keegan, 2006; Gomez & Wilson, 2006; Henry & Sears, 2002; Sidanius & Pratto, 1999; Sommers & Norton, 2011). For example, Kinder and Sears (1981) found that symbolic concerns and resentment about racial progress were better predictors of voting behavior and public policy attitudes than the threat (measured both subjectively and objectively) blacks posed to whites, including desegregation, economic competition, and crime.

White Americans' resentment towards racial progress, and their attitudes about existing inequality and public policy intended to improve the welfare of racial minorities, amounts to a justification for and legitimization of pervasive racial inequality and discrimination. Some race scholars have referred to these social conditions as "laissez fair racism", in which White Americans view racial minorities as the "architects of their own disadvantaged status" (Bobo, Kluegel, & Smith, 1996, p. 22). From this perspective, overtly hostile and inegalitarian racial attitudes are no longer necessary for the maintenance of racial hierarchy. Instead, White Americans' unwillingness or inability to recognize, acknowledge, or confront patterns of racial discrimination and inequality is sufficient for its preservation.

Attributions for the causes of racial disadvantage may be fundamental to the persistence of white's ignorance of or apathy for the plight of racial minorities, despite their explicit egalitarian preferences. Indeed, racial stereotypes and belief in the inferiority of black culture may provide a persuasive justification for prejudicial attitudes and resulting discriminatory outcomes. For example, negative racial stereotypes are particularly relevant to whites' attitudes towards social welfare and criminal justice policies (Huddy & Feldman, 2009; Sniderman & Piazza, 1993). The belief that welfare

recipients are predominantly black and that black people lack work ethic can explain much variability in whites' opposition to public assistance (Gilens, 1999). Similarly, stereotypes about blacks as criminal, violent, and aggressive predict white's perception that discriminatory outcomes in the criminal justice system are justified and fair. These personal attributions for unfair treatment and discrimination experienced by blacks also increase support for more punitive criminal justice policies among whites (Peffley & Hurwitz, 2010).

In short, White Americans' use of racial stereotypes to justify or outright ignore racial discrimination and inequality help preserve a belief in the legitimacy of existing social hierarchy and the status quo. Accompanying, if not motivating, this attitudinal orientation on race relations among White Americans is a false belief in egalitarian social arrangements. Social psychologists have long recognized that social hierarchy and inequality is a common feature of human social arrangements, and that people adopt belief systems that allow them to justify and perceive as legitimate existing social hierarchy. For example, according to social dominance theory (Pratto et al., 1994; 2006), societies minimize intergroup conflict and justify the legitimacy of social hierarchy and inequality through the construction of mutually shared beliefs and ideologies. Examples of hierarchy-legitimizing myths include meritocracy, political-economic conservatism, and nationalism (Pratto, Sidanius, & Levin, 2006). Egalitariansim, as a hierarchy-legitimizing myth, can also be understood as a socially constructed justification for the existence of group-based dominance, which makes acceptable out-group derogation (e.g., stereotypes), in-group favoritism and, consequently, the perpetuation of the status quo and existing social hierarchy.

The status quo is a powerful psychological anchor; people are motivated to believe that current social conditions and systems are good, fair, and just (Kay et al., 2008). Much empirical evidence suggests that threats to the legitimacy of the status quo activates system justification motives, particularly among individuals most invested in its maintenance (e.g., Jost, Banaji, Nosek, 2004). Specifically, the activation of system justification motives by the threat to the legitimacy of the status quo increases individuals' endorsement of system-justifying ideologies and stereotypes, and may lead individuals to view current social condition and arrangements (including group-based inequalities) as desirable (i.e., "injunctification"; Kay, Gaucher, Napier, Callan, & Laurin, 2010). Evidence of racial inequality and discrimination may fundamentally threaten individuals' belief in the legitimacy and fairness of the status quo, and, consequently, increase the motivation to justify or rationalize existing racial hierarchy and engage in injunctification.

White Americans tendency to justify and legitimize (or altogether ignore) racial inequality while condemning discrimination signifies the emergence of a new form of explicit racial prejudice that replaces the more overtly hostile forms common in a bygone era. More importantly for the current work, this new form of racial prejudice raises major questions about the ability for laypeople to understand and adjust for the existence and consequences of their own implicit racial bias.

Implicit Racial Bias and Explicit Egalitarianism

This discrepancy between racially tolerant values and the acceptance of existing racial disparities indeed indicates the emergence of a new form of racial prejudice. Of central importance to this subtle form of racial prejudice is the role of egalitarian social

norms in attenuating the expression of hostile prejudice, and the persistence of unconscious racial biases that have long influenced judgment and behavior in ways that may now be inconsistent with consciously accessible, egalitarian beliefs. In this section, I review evidence for the existence of implicit racial biases and its implications for discriminatory judgment and behavior that may contribute to aggregate-level racial disparities.

Egalitarian social norms have increased White Americans' motivation to *appear* unprejudiced and unbiased in their interactions with racial minorities (e.g., Bersieker, Shelton, & Richeson, 2010; Norton, Sommers, Apfelbaum, Pura, & Ariely, 2006; Plant & Butz, 2006; Shelton, 2003; Vorauer, Hunter, Main, & Roy, 2000; Vorauer, Main, & O'Connell, 1998). Most people are largely aware that expressions of racial prejudice are considered socially unacceptable (Crosby, Bromley, & Saxe, 1980; Monteith, Deneen, & Tooman, 1996). Thus, one explanation for the persistence of racial disparities despite the rise of egalitarian social norms is that respondents are unwilling or unable to accurately report their personal beliefs about racial minorities. Researchers investigating attitudes about race relations with self-report measures may not be able to accurately capture the attitudinal construct, but instead promote respondents' acquiescence to perceived situational demands and norms in socially desirable ways.

However, concerns about the potential for response bias to undermine the reliability and validity of self-reported measures of racial attitudes are an incomplete explanation for the persistence of racial discrimination and inequality. For example, White Americans' are not only motivated to appear unprejudiced, but are also motivated to minimize the impact of racial prejudice on their judgment and behavior (e.g., Plant &

Devine, 1998). Furthermore, many White Americans have internalized egalitarian standards for their personal conduct that regulates the expression of prejudicial attitudes (Moneith, 1993). Later, I discuss the implications of this motivation to control prejudice and the internalization of egalitarian social norms for the acceptance of evidence and allegations of implicit racial bias. For now, it is important to note that, because of the explicit belief in egalitarian values and the motivation to adhere to egalitarian principles, many scholars have suggested that unconscious, implicit forms of bias at the individual-level may help explain aggregate racial disparities (e.g., Banaji & Greenwald, 2013).

Indeed, a substantial body of research suggests that the majority of White Americans can be characterized by relatively negative unconscious associations towards racial minorities and other stigmatized groups that is nonetheless consequential for social judgment and behavior (Blair, 2001; Greenwald & Banaji, 1995; Plant, Devine, & Brazy, 2003). Attitudes and other associations can be represented at both the implicit and explicit levels. Explicit associations are consciously accessible, and can be retrieved and reported with accuracy, whereas implicit associations may not be consciously accessible and can operate quickly without conscious awareness, effort or control (Bargh, 1994; Gawronski et al., 2006). Due to these differences, implicit associations cannot be measured using the direct measures that are used to assess explicit associations, like surveys, and are instead assessed using latency-based, indirect measures (see De Houwer, 2006). These indirect measures of implicit attitudes are much less susceptible to respondent misrepresentation than the direct measures of explicit attitudes, at least among participants who are not explicitly instructed on how to fake the test (Banse, Seise, & Zerbes 2001; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Indirect measures may

also be better predictors of respondents' behavior than self-reported measures in socially sensitive domains (Gawronski et al., 2006**;** Greenwald et al., 2009), or when individuals lack conscious awareness of or access to their attitudinal preferences (e.g., undecided voters; Hawkins & Nosek, 2012; Arcuri et al., 2008).

Psychological phenomena that are operating at the unconscious level or are otherwise inaccessible to conscious awareness, including the activation of and reliance upon stereotypes, social categories, schemas, heuristics, and prior expectancies, can be understood as automatic mechanisms that can influence social perception, judgment, and behavior in profound ways (Baron & Banaji 2006; Bargh, 1994; Dasgupta et al., 2000 Greenwald & Banaji, 1995; Morsella & Bargh, 2011). Implicit racial biases have been found to predict, using a broad range of measurement strategies, discriminatory attitudes and behavior in a wide range of laboratory and field contexts, including general evaluations of racial minority targets (Amodio & Devine, 2006), mock juror decision-making (Levinson, Cai, & Young, 2010), employment and promotion decisions (Scheck, 2004; Zeigert & Hanges, 2005), medical consultation (Penner et al., 2012; Green et al., 2007), citizens' and police officers' decision to fire a weapon at an unarmed suspect (Plant & Peruche, 2005; Sadler et al., 2012), support for public policy intended to help racial minorities (Rudman & Lee, 2002), and voting in the 2008 U.S. Presidential election (Greenwald, Smith, Sriram, Bar-Anan, & Nosek, 2009), among others (see Jost et al., 2009 for a review).

More importantly, because these implicit associations are generally not subject to conscious scrutiny or impression-management strategies (Nosek, 2005), individuals' self-reported, explicit attitudinal preferences or values may be in direct conflict with their

unconscious associations. As demonstrated in several meta-analyses (Greenwald et al., 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) and other studies (e.g., Nosek, 2005; Nosek & Smyth, 2007), implicit and explicit measures of attitudes and other associations both show incremental predictive validity across behavioral domains. This finding suggests that each attitudinal construct may be measuring distinct psychological processes or cognitive structures that account for distinct portions of the variance in attitudinal and behavioral outcomes.

Because unconscious racial biases can lead to discriminatory judgment and behavior in ways that may be inconsistent with explicit egalitarian intentions and ideals in contemporary times, it may also compromise White Americans' ability to maintain and project unprejudiced self-images. Thus, many people are threatened by information, situations and interactions that might threaten their egalitarian self-images, which has important implications for the conditions under which White Americans will be receptive to feedback on their level of implicit racial bias.

Unprejudiced Self-Image Maintenance

In the previous section, I argued that the motivation to appear unprejudiced may undermine the validity of self-report measures of racial attitudes, but remains only a partial explanation for the persistence of racial discrimination and inequality in an egalitarian society. While explicit sociopolitical attitudinal orientations may lead many White Americans to rationalize racial inequality in socially acceptable terms, unconscious yet consequential forms of racial bias may *directly* contribute to patterns of racial discrimination, despite individuals' genuine egalitarian intentions and beliefs (e.g., Dovidio & Gaertner, 2000, 2004; Zeigert & Hanges, 2005). In this section, I argue that

egalitarian social norms have also increased the social and moral costs of being prejudiced, which heightens the threat associated with attributions of racist intent, increases the psychological need to maintain an unprejudiced self-image, but leaves the psychological legacy of racial bias largely in place.

Few social labels are more consequential or aversive than being categorized as a racist (Crandall, Eshelman, & O'Brien, 2002). Indeed, White Americans' are aware of and threatened by the possibility of being stereotyped as a racist (Frantz et al., 2004). More generally, prior work suggests that people often feel threatened by the prospect of confirming a negative stereotype about their social group. This stereotype threat can undermine performance in ways that confirm the stereotype, particularly among people who most strongly identify with that domain (Steele, 1997). For example, Frantz et al. (2004) find that White Americans who most strongly identify as egalitarian (indexed by the motivation to control prejudice; Dunton & Fazio, 1997) demonstrated higher levels of implicit racial bias when the threat of appearing racist was salient (e.g., completing a test diagnostic of racial prejudice).

Thus, many people are threatened by information and/or interactions that might threaten their egalitarian self-images, and are motivated to avoid the attribution of racist intent. The threat of being negatively stereotyped as a racist may be particularly salient for people who consciously endorse egalitarian beliefs and sincerely believe themselves to be unprejudiced, but who nonetheless harbor negative attitudes and beliefs towards racial minorities outside their conscious awareness. Several prominent perspectives on contemporary racial prejudice focus on the conflict between egalitarian social norms and personal standards, and underlying unconscious negative feelings and beliefs towards

racial minorities (e.g., McConahay, 1986; Sears, Henry, & Kosterman, 2000).

For example, the theory of aversive racism (e.g., Gaertner & Dovidio, 1986, 2000) characterizes an ambivalent attitudinal orientation among White Americans towards racial minorities. This ambivalence arises between egalitarian ideals and unconscious racial biases. In contrast with explicit, traditional racists, aversive racists consciously endorse egalitarian norms, and genuinely regard themselves as unprejudiced, yet harbor unconscious racial biases. In ambiguous situations or when discrimination can otherwise be expressed in subtle or socially acceptable terms, unconscious biases among aversive racists can lead to patterns of discrimination that disadvantage racial minorities (e.g., Zeigert & Hanges, 2005). As a result, aversive racists are particularly threatened by information that challenges their egalitarian self-images, and are avoidant of situations, interactions, and behaviors that may allow for an attribution of racist intent.

Avoidance of or aversion to situations and interactions that may threaten one's egalitarian self-concept and social standing is certainly one way in which White Americans maintain an unprejudiced self-image. Additional research suggests that people also engage in a range of social and cognitive strategies to protect the egalitarian self and avoid the attribution of racist intent (e.g., O'Brien et al., 2010; Sommers & Norton, 2006). These strategies for persuading both the self and others of one's egalitarianism, however, suggest that many White Americans are unable or unwilling to recognize the existence of racial discrimination, let alone implicit racial bias and its consequences for racial minorities. For example, strict adherence to a "color-blind" ideology communicates publically a belief in egalitarian principles and opposition to race-based decision-making. However, doing so also indirectly denies the unique identity and cultural history of racial

minorities (Norton, Vandello, Biga, & Darley, 2008), belies an increased tendency to express both implicit and explicit biases (Richeson & Nussbaum, 2004), or otherwise serves as a justification for existing inequality (Peffley & Hurwitz, 2010).

Furthermore, many White Americans engage in self-serving social comparisons with the stereotypical 'bigot' to protect the unbiased self-image (O'Brien et al., 2010). That is, people often obtain and confirm knowledge about the self via comparisons with relevant others (Festinger, 1954). However, social and cultural representations of racial prejudice largely anchor on explicit, hostile or more traditional expressions of racial animosity (Vera, Feagin, & Gordon, 1995; Feagin & Vera, 1995; Sommers & Norton, 2006). Downward social comparisons with the stereotypical representation of the racist bigot allow for the inference that one is not, in fact, a full-blown racist, and may reduce one's effort to regulate the expression of their bias (O'Brien et al., 2010). More generally, lay perceptions of what behavior and traits characterize a bigot and constitute racial prejudice can function to distance oneself from the category 'racist' and to maintain the unprejudiced self-image. For example, White Americans are less likely to describe subtle, ambiguous forms of bias as indicative of racial prejudice (Sommers & Norton, 2006). Unintentional, unconscious bias does not align with existing cognitive schemas of the traditional White bigot.

Together, these findings emphasize that most White Americans may be particularly threatened, and motivated to deny or rationalize, evidence of their own racial bias as not constituting racial prejudice. White Americans certainly recognize the social consequences of appearing racist and many also recognize that they harbor some form of racial prejudice (Saucier, 2002). However, this level of social and personal awareness

may not necessarily translate into the inference that one's bias is problematic, consequential and therefore in need of control. Indeed, it may be that individuals most accurately characterized by racial prejudice are also least able to recognize it in themselves (Devine & Monteith, 1999; Sommers & Norton, 2006; Monteith & Mark, 2009).

The motivated rejection of information that threatens the egalitarian self-image may be particularly true for subtle forms of racial bias operating outside of conscious awareness. The existence of implicit racial bias is inconsistent with lay beliefs of racial prejudice and cultural representations of racial bigotry. Furthermore, the expression of implicit racial bias is often more ambiguous, absent of conscious intent, and is more easily justified or attributed to socially acceptable causes. Consequently, it is likely that many White Americans will respond to allegations of racial bias on the basis of subtle, ambiguous, implicit attitudes with hostility, denial, and defensiveness (Howell et al., 2013; Howell et al., 2014; Kaiser & Miller, 2001). Nevertheless, additional work on the regulation of prejudice suggests that despite this concern, individuals who are genuinely and intrinsically committed to being nonprejudiced may be more accepting of evidence of their own racial bias (Devine et al., 1991; Monteith & Voils, 1998; Zuwerink et al., 1996). Instead of responding with defensiveness and denial, when these individuals perceive that their own behavior is inconsistent with their egalitarian standards, they become more motivated to regulate their prejudice. Below, I review the literature on the motivation to control and regulate prejudice responding.

Motivation to Control and Regulate Prejudiced Responding

In the previous sections, I argued that egalitarian social norms increases the social

and moral costs of appearing and being prejudiced, which heightens the threat of being labeled a 'racist', undermines the validity of self-report measures, and increases the psychological need to maintain an unprejudiced self-image. Consequently, a broad range of cognitive strategies and outright behavioral avoidance are commonly deployed to protect the unprejudiced self-image. Thus, many White Americans will respond defensively to evidence and allegations of implicit racial bias.

It is important to note, however, that egalitarian social norms have *also* promoted the motivation to control for prejudice responding and to minimize its impact on social judgment and behavior (Bergsieker, Shelton, & Richeson, 2010; Dunton & Fazio, 1997; Glaser & Knowles, 2008; Monteith et al. 1993; Monteith & Mark, 2009; Plant & Devine 1998). Indeed, much evidence indicates that White Americans' vary in their motivation to control expressions of prejudice, and prejudice-regulation is difficult, time-consuming, and cognitively taxing (Devine, 1989; Devine & Monteith, 1993; Richeson & Shelton, 2003). Nonetheless, the motivation to control prejudice responding has been found to moderate the expression of both explicit and implicit racial biases (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Frantz et al., 2004; Glaser & Knowles, 2008; Zeigert & Hanges, 2005). Thus, activating this motivation may be central to the efficacy of prejudice regulation and anti-bias interventions (Monteith & Mark, 2009).

The self-regulation of prejudice varies in the extent to which it is motivated by intrinsic, personal, or self-determined reasons (Devine et al., 2002; Legault, Green-Demers, Grant, & Chung, 2007; Legault et al., 2007; Legault, Gutsell, & Inzlicht, 2011; Plant & Devine, 1998). Whereas intrinsic motivations to respond without prejudice arise from internalized valuation of egalitarianism, extrinsic motivations to regulate prejudice

may derive, instead, from the desire to avoid the attribution of racist intent and the stigma of being labeled a racist. In this regard, the egalitarianism of White Americans' who are motivated to control prejudice for extrinsic reasons can be understood as a less sincere impression-management strategy, which is unlikely to adequately regulate the expression of bias (Amodio, Harmon-Jones, & Devine, 2003; Plant & Devine, 1998). Consistent with this idea, people who are motivated to regulate prejudice for external reasons (e.g., social norms), but not internal (e.g., personal beliefs) reasons, express higher levels of racial bias and evidence greater reactance, defensiveness, and resentment when pressured to comply with egalitarian norms (Amodio, Harmon-Jones, & Devine, 2003; Devine et al., 2002; Legault, Green-Demers, & Eadie, 2009; Legault et al., 2007; Plant & Devine, 2001; Plant, Devine, & Peruche, 2010). These findings suggest that White Americans motivated to control their prejudiced responses primarily for extrinsic reasons will be particularly motivated to reject evidence and allegations of implicit racial bias.

In contrast, many White Americans have internalized egalitarian standards for their own personal conduct, which has been found to reduce the expression of racial prejudice, particularly among individuals characterized by low levels of racial bias (Devine et al. 2002; Legault et al., 2011). That is, individuals characterized by low levels of racial prejudice view egalitarianism as an important component of their self-concepts (Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1993; Monteith, Devine, & Zuwerink, 1993; Zuwerink, Monteith, Devine, & Cook, 1996). These personal norms (i.e., Schwartz, 1977) can be understood as self-imposed expectations for one's behavior. However, egalitarian standards of conduct may also develop based on the influence of important others and perceived social norms. For example, the egalitarian standards for

high prejudiced individuals are not as closely assimilated into their self-concepts and are less related to a sense of moral obligation (Monteith, Deneen, Tooman, 1996). Instead, egalitarian standards for high prejudiced individuals are largely based on the standards imposed by others (Monteith et al., 1993[2]). Thus, both the motivation to control prejudice responding and the acceptance of egalitarian standards for personal conduct each vary in the extent to which it reflects intrinsic, personal reasons compared to extrinsic, socially imposed expectations.

The distinction between intrinsic and extrinsic motivation to control prejudice and internalize egalitarian standards for personal conduct may have important implications for how White Americans respond to prejudiced-related discrepancies, or evidence that one has acted in ways inconsistent with egalitarian standards. People at all levels of prejudice often report that they sometimes respond to racial minorities in ways that are inconsistent with their egalitarian ideals. Even the majority of low-prejudiced individuals report that they experience racially biased thoughts and feelings, and many often report difficulty controlling prejudice-responding (Devine, Monteith, Zuwerink, & Elliott, 1991; Monteith, Devine, & Zuwerink, 1993; Monteith, 1993; Monteith, Voils, Ashburn-Nardo, 2001). Stereotypes about racial minorities can be activated and applied to racial minorities automatically and without conscious intent for most White Americans (Banaji et al., 1993; Banaji & Greenwald, 1995; Devine, 1989; for exceptions, see Fazio et al.,

[2] A major exception to this general pattern is the work by Monteith & Walters (1998), who demonstrate that high prejudiced individuals who operationalize egalitarianism in term of "equality of opportunity" do, in fact, feel morally obligated to regulate prejudice responding. However, most work supports the idea that highly prejudiced, externally motivated individuals do not feel morally obligated to be unprejudiced. I do not distinguish between different operationalizations of egalitarianism in my theory or proposed measures.

1995). And, as reviewed above, much evidence (e.g., Greenwald & Banaji, 1995; Plant, Devine, & Brazy, 2003) indicates that the majority of White Americans can be characterized by unconscious racial bias that can influence their judgments and behavior in ways inconsistent with their explicit beliefs and values. Despite their conscious intentions and egalitarian standards, most White Americans experience some degree of prejudice-related discrepancies, regardless of their level of expressed racial prejudice (Monteith, 1996).

More importantly, the affective reactions and behavioral consequences of prejudiced-related discrepancies (i.e., perceived violations of one's egalitarian standards) vary across individuals' expressed level of prejudice and their intrinsic motivation to control for it (e.g., Monteith, 1993). Among individuals characterized by low levels of prejudice and the intrinsic motivation to control prejudice responding, prejudice-related discrepancies lead to feelings of compunction, shame and uneasiness (Devine et al., 1991; Monteith & Voils, 1998; Zuwerink et al., 1996). These feelings of guilt due to prejudice-related discrepancies among low-prejudiced individuals serves as a cue for the need to adjust for one's bias and attenuate the expression of prejudice (Devine et al., 2002). In contrast, among individuals characterized by high levels of prejudice and the extrinsic motivation to appear egalitarian, prejudiced-related discrepancies engender other-directed negative affect, such as resentment towards the stigmatized group (Monteith, 1993; Strauman & Higgins, 1987). This resentment, although triggered by the awareness of one's own racial bias, can be understood as a self-protective orientation that undermines the motivation to regulate prejudice and reduce its impact on judgment and behavior.

In short, low prejudice individuals may still experience the automatic activation and application of negative stereotypes and evaluations towards racial minorities outside of conscious awareness (Monteith, Voils, & Ashuburn-Nardo, 2001). However, these individuals are distinguished from their highly prejudiced peers by becoming highly motivated to attenuate the expression of bias. These efforts to reaffirm one's egalitarianism result from the experience of negative self-directed affect when low prejudice individuals perceive that their conduct has violated internalized, intrinsically motivated egalitarian standards (Monteith, 1993; Monteith, Spicer, & Tooman, 1998; Monteith & Walters, 1995). Nevertheless, it remains unclear and untested if evidence of one's unconscious racial bias will lead to the same affective and behavioral consequences among low prejudiced, intrins egalitarians as has been found for other forms of prejudice-related discrepancies that are more commonly studied in the literature.

More specifically**,** at the core of research and theory on prejudice-regulation is the role of self-awareness that one's past or future behavior may be inconsistent with egalitarian standards (Monteith & Mark, 2009). For example, most research focusing on the self-regulation of prejudice has experimentally induced the perception that one's behavior has violated egalitarian ideals (e.g., Monteith et al., 1993) or has measured the self-reported tendency towards prejudice-related discrepancies across hypothetical situations (e.g., should-would discrepancies; Monteith, Devine, & Zuwerink, 1993; Monteith & Voils, 1998). This work has not directly focused on the consequences of prejudice-related discrepancies due to awareness of one's implicit racial bias. Furthermore, prior research has not considered the antecedents and consequences of

individuals' unwillingness or inability to recognize the existence of unconscious racial biases.

The current program of research seeks to address these gaps in existing research and theory on aversive racism, prejudice regulation, and egalitarian self-image maintenance. In particular, the proposed research is the first to determine whether all self-identified egalitarians respond defensively to information that threatens their egalitarian self-image, or if, instead, only intrinsically motivated egalitarians respond to evidence of their prejudice-related discrepancies with an increased acceptance of implicit racial bias and the motivation to regulate prejudice.

Current Research: Overview of Theory, Hypotheses, and Methods

Despite contemporary egalitarian social norms, values, and personal standards, pervasive racial inequality exists. White Americans' explicit racial and sociopolitical attitudes render them incapable or unwilling to recognize patterns of racial discrimination, or support public policies intended to improve the lives of racial minorities. However, this attitudinal orientation cannot clearly be understood in terms of hostile, overt racial bigotry. Most White Americans' sincerely believe that they should not and do not hold prejudicial attitudes or beliefs towards racial minorities. Instead, contemporary forms of racial prejudice represent a more subtle racial bias and apathy that can be rationalized in socially acceptable terms.

Sociopolitical Orientations and Perceptions of Racial Discrimination

Any evidence of racial bias, inequality and discrimination, to the extent that it threatens the perceived legitimacy of the status quo and existing social hierarchy, is likely to motivate some White Americans to become even more entrenched in their

endorsement of racial stereotypes, rationalization of inequality, and opposition to racially liberal public policy. For these reasons, I expect that people characterized by explicitly hostile racial attitudes, resentment towards racial progress, system justification, or social dominance orientation will be motivated to reject evidence of aggregate-level racial prejudice and discrimination, as well as feedback suggesting that they harbor implicit racial bias (Hypothesis 1). Similarly, any credible evidence of racial discrimination and inequality, and feedback suggesting that one harbors implicit racial bias, will strengthen or activate the effect of racial resentment, system justification, and social dominance orientation on racial stereotyping and the motivated rejection of all evidence of racial discrimination and inequality and implicit racial bias (Hypothesis 2). Evidence for motivated reasoning processes will be indexed by 1) self-perceived bias, 2) belief in the existence and consequences of implicit racial bias, 3) perceptions of the credibility, objectivity and validity of measures of implicit attitudinal constructs, and 4) attitudes towards social science and scientists.

Intrinsic vs. Extrinsic Egalitarianism and Prejudice-Related Discrepancies

Of central importance to the current investigation, however, is the role of egalitarian norms, ideals, and personal standards in shaping reactions to and acceptance of evidence of one's own implicit racial bias, which has not been examined in prior research. In general, I expect that people characterized by an extrinsic motivation to be egalitarian will be unwilling to acknowledge the general existence of implicit racial bias. In contrast, White Americans' with an intrinsic motivation to regulate prejudice and who have internalized egalitarian standards for their personal conduct will be most receptive to evidence documenting the existence of implicit racial bias in the population.

Acceptance of implicit bias in the general population among intrinsic egalitarians should be particularly likely when these individuals encounter credible evidence to support that inference (Hypothesis 3). Still, it remains unclear if this receptivity to evidence of implicit racial bias in general also translates to acceptance of allegations of one's own bias, or if instead it activates motivated reasoning processes that lead to a more self-protective orientation against threats to one's egalitarian self-image. I have reviewed evidence that supports both predictions that 1) White Americans' are particularly threatened by and avoidant of information that challenges their unprejudiced self-image (e.g., Frantz et al., 2004; Gaertner & Dovidio, 1986, 2000; Spencer et al., 1998), and 2) that egalitarians respond to prejudice-related discrepancy with compunction and increased efforts at prejudice-regulation (e.g., Monteith & Mark, 2009).

Because egalitarian social norms have increased the social and moral costs of appearing and being prejudiced, many people are threatened by information and social interactions that might challenge their egalitarian self-concepts. Unfortunately, contemporary forms of racial prejudice can bias judgment and behavior outside of conscious awareness, and in ways that may directly conflict with even sincerely held egalitarian values. Indeed, many White Americans can be characterized by this kind of an ambivalent attitudinal orientation towards racial minorities (Gaertner & Dovidio, 1986).

To protect the unprejudiced self-image, people often engage in a range of cognitive strategies, including self-serving social comparisons, endorsement of "color-blind" ideologies, and holding narrow definitions of what constitutes racial prejudice. These strategies for persuading the self and others of one's egalitarianism allows for the rationalization of socially unacceptable manifestations of racial prejudice, particularly its

implicit forms. Many White Americans are simply avoidant of and averse to situations and interactions that may lead to attributions of racist intent.

More generally, people are motivated to maintain congruency between the set of expectations and meanings incorporated into their self-concepts, and appraisals (by the self and others) of their behavior (Burke, 2006; Stryker & Burke, 2000; Swann, 1987, 2005; Swann & Bosson, 2008). Incongruency is experienced as distressing (Burke, 1991, 2006; Burke & Harrod, 2005; Cast & Burke, 2002). One psychological consequence of such distress is the heightened motivation to reestablish congruency by distancing oneself from the source of incongruency or by otherwise resisting its implications for perceptions of the self. Indeed, people often derogate and dismiss information that threatens a desired self-image (e.g., Shepperd, Malone, & Sweeny, 2008; Spencer et al., 1998). Together, this work suggests that, in general, all White Americans will be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias. Similarly, White Americans with dissociated implicit and explicit racial attitudes- e.g., low explicit racial biases, high implicit racial biases- should also be motivated to reject evidence of their own implicit racial biases

However, prior research on prejudice-regulation and the motivation to control prejudiced responding strongly points to the possibility that not all self-identified egalitarians will respond to allegations of racial biases with defensiveness and outright rejection. In particular, prior research suggests that egalitarian values and prejudice regulation varies in the extent to which it represents internalized, intrinsic motivations (e.g., Monteith, 1993). White Americans who are motivated to control their prejudiced responses primarily for intrinsic reasons might experience self-directed negative affect

upon learning that they harbor unconscious racial bias. From this perspective, increased feelings of guilt and shame among intrinsic (vs. extrinsic) egalitarians should attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation (Hypothesis 4a)**.** Indeed, this pattern of results has been observed for intrinsic egalitarians who report experiencing prejudice-related discrepancies or received feedback suggesting that their behavior was discriminatory (e.g., Monteith et al., 1993). In contrast, extrinsic egalitarians will be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias (Hypothesis 4b). Thus, it is also expected that intrinsic (vs. extrinsic) egalitarians with dissociated implicit-explicit racial attitudes (e.g., low explicit prejudice, high implicit prejudice) will be more accepting of evidence of their own implicit racial bias (Hypothesis 5).

However, there are several reasons to expect even intrinsic egalitarians to be discomfitted and therefore motivated to *reject* direct evidence of their own implicit bias, instead of responding to this information with a motivation to regulate prejudice and affirm their egalitarianism (e.g., Monteith, 1993). For example, because implicit attitudes are activated automatically and without conscious awareness, its influence on social judgment is difficult to control or change (e.g., Gregg, Seibt, & Banaji, 2006), let alone recognize. People may perceive a lack of self-efficacy to control the expression of unconscious bias or attenuate its impact on judgment and behavior. A lack of self-efficacy has been found to lead to defensive posturing and behavioral avoidance in response to other kinds of self-threatening information designed to encourage behavioral change (e.g., health domain; Witte & Allen, 2000).

Further, unlike traditional forms of racial prejudice, contemporary forms of subtle, unconscious racial bias violate lay peoples' intuitions and beliefs about how racial prejudice is expressed and by whom **(**Sommers & Norton, 2006**).** The existence of one's implicit racial bias is likely to be inconsistent with consciously accessible egalitarian beliefs and attitudes. This kind of self-knowledge may be a more subjectively relevant standard of comparison for evaluating the appropriateness of one's attitudes or behavior than external feedback. Consequently, evidence of one's implicit racial bias may not activate a perceived discrepancy between personal standards and behavior in a way that is sufficient to motivate prejudice regulation. Indeed, self-reported prejudice-related discrepancies do not appear to be related to implicit measures of racial bias (Monteith, Voils & Ashburn-Nardo, 2001).

Consistent with this line of reasoning, one study used implicit measurement techniques to study the detection of implicit racial bias and its relation to prejudice regulation (i.e., Monteith, Voils & Ashburn-Nardo, 2001). Results indicated that most individuals were able to detect their biased performance on the IAT (e.g., slower response times to stereotype-incongruent trials). When biased performance on the implicit association test was attributed to racial prejudice, participants experienced the type of self-directed negative affect that has traditionally been found to promote prejudice regulation. However, most participants misattributed biased performance to nonracial factors. This was particularly true among individuals' who reported low levels of prejudice-related discrepancies. These findings suggest that even White Americans who are relatively successful at prejudice regulation (e.g., intrinsically motivated, low-prejudiced egalitarians) may nonetheless fail to detect their own implicit, unconscious

racial biases or attribute it to prejudice responding. Without self-awareness of one's own implicit racial bias, prejudiced-regulation is unlikely to be activated (e.g., Czopp, Monteith & Mark, 2006). Still, it remains possible, albeit untested, that prejudice-regulation will be activated among low prejudiced, intrinsic egalitarians who are *directly* confronted with information suggesting that their biased performance on the IAT does, in fact, indicate racial prejudice (Hypothesis 4b).

In contrast, extrinsically motivated egalitarians (i.e., high-prejudiced individuals whose egalitarian standards are based on expectations imposed by others) tend to respond to prejudice-related discrepancies with resentment, and a decreased motivation to regulate prejudice (e.g., Monteith, 1993). Threats to a socially desired self-image has been found to reduce self-esteem (Baumeister & Tice, 1985; Rudman, Dohn, & Fairchild, 2007) and promote out-group derogation and stereotyping (Colange, Fiske, & Sanitioso, 2009; Spencer et al., 1998; Frantz et al., 2004); Fein & Spencer, 1997). Thus, I expect high prejudiced individuals who are motivated to project an egalitarian self-image primarily for extrinsic reasons or in compliance with socially imposed norms and expectations to be motivated to reject evidence of their own unconscious racial bias (Hypothesis 6).

Downstream Consequences of the Motivated Rejection of Implicit Bias

Importantly, the motivated rejection of evidence and allegations of implicit bias and its relationship to aggregate racial disparities have several important downstream consequences not considered by prior research. For example, the motivated rejection of implicit bias feedback is expected to increase personal (vs. situational) attributions for the causes of aggregate-level racial disparities (Hypothesis 7), which in turn should increase

support for more punitive attitudes towards the criminal justice system (Hypothesis 8), and opposition to policy remediation to address racial inequality such as affirmative action and social welfare (Hypothesis 9). Furthermore, the effects of these motivated reasoning processes are also expected to undermine individuals' perceived value and importance of organizational interventions designed to reduce implicit racial bias (Hypothesis 10), and the general motivation to control prejudiced responding and its impact on behavior (Hypothesis 11; Dunton & Fazio, 1997).

Interventions to Attenuate Motivated Rejection of Implicit Bias

Policymakers and organizations spend billions of dollars annually on prejudice interventions (Hansen, 2003). Many of these proposed remedies lack empirical evidence (Paluck & Green, 2009). However, some anti-prejudice strategies targeting implicit attitudes that have the most empirical support commonly involve improving White Americans' *awareness* of implicit cognition (e.g., Blair, 2002; Banaji, 2001; Devine et al., 2012; Paluck & Green, 2009), highlighting the potential discrepancy between their implicit and explicit attitudes (Correll et al., 2002; Monteith et al., 2002), or identifying strategies to overcome their implicit biases (Devine et al., 2012; Lai et al., 2015). Providing individuals with accurate information about their implicit racial biases may therefore be a valuable approach to prejudice-reduction interventions (e.g., Hillard, Ryan & Gervais, 2013) and for encouraging prejudice-regulation (e.g., Monteith et al., 1993). But, if White Americans' characterized by implicit racial bias are unwilling or unable to recognize it, then anti-prejudice interventions will fail to reduce the impact of racial attitudes on judgment and behavior (Sommers & Norton, 2006). Anti-prejudice interventions that improve both social and introspective awareness of implicit racial bias

without engendering resentment and motivated reasoning processes are likely to be among the most effective.

Accordingly, the current research also tests two independent interventions that are expected to reduce the motivated rejection of allegations of one's own implicit racial bias and its downstream effects described above (Hypothesis 12). The first intervention is designed to communicate the descriptive norm that implicit racial bias is common in the general population, and is indeed a basic feature of human cognition. Evidence of one's own implicit racial bias may be experienced as a type of moral condemnation uniquely reserved for the self- a personal failure to meet contemporary egalitarian social norms, and an implied complicity in the perpetuation of a long history of racial bigotry. Emphasizing the commonality of implicit bias in the general population is therefore expected to be a fruitful avenue for reducing the motivated reasoning processes activated by the threat posed by one's own implicit racial bias. The underlying logic to this intervention is that this descriptive norm will diffuse the threat associated with evidence of one's implicit racial bias, and potentially disarm the belief that one is unique for harboring unconscious bias.

However, one major concern with this intervention is the risk of communicating the normative acceptability of implicit forms of racial bias and undermining individuals' motivation to adjust for its influence on judgment and behavior. Thus, by increasing the belief that implicit racial bias is the norm, people may infer that such bias is not in need of change. As a result, this intervention could lead to ironic effects by undermining prejudice-regulation motivation and the perceived value of anti-prejudice interventions (Duguid & Thomas-Hunt, 2014). Nevertheless, communicating descriptive social norms

about implicit racial bias is an intuitive interpersonal strategy for promoting awareness of implicit cognition, and is a common component of many anti-prejudice interventions.

The second intervention is designed to mitigate the threat to the unprejudiced self-image by bolstering the worth and integrity of the self prior to feedback about one's implicit racial bias.

According to self-affirmation theory (Steele, 1988), individuals are often motivated to resist information that threatens the perceived worth and integrity of the self. Such defensive biases in response to personally threatening information can function to preserve the positivity of individuals' self-perceptions. A substantial body of research indicates that affirming the global adequacy of the self prior to encountering information or situations that threaten the positivity of one's self-concept can reduce defensiveness and biased responding (e.g., McQueen and Klein, 2006; Sherman & Cohen, 2002; Sherman, Nelson, & Steele, 2000). In the context of intergroup attitudes, for example, self-affirmation has been found to increase White Americans' awareness of discrimination towards racial minorities (Adams, Tormala, & O'Brien, 2006), reduce the expression of racial prejudice (e.g., Fein & Spencer, 1997), and attenuate the threat of being perceived as racist (Frantz et al., 2004). Along these lines, it is expected that allowing White Americans to affirm the integrity and positivity of their self-concept prior to receiving feedback about their implicit racial bias will attenuate motivated reasoning processes.

Overview of Studies 1, 2, and 3

In my thesis research, I conducted three experiments to systematically test these hypotheses (see Appendix for full list of predictions, experimental materials, and

measures). Study 1 was designed to manipulate the belief in the existence of implicit bias by providing participants with credible information that summarizes psychological research and theory on implicit cognition and racial discrimination. Study 2 experimentally manipulated participants' belief in their own implicit bias using a false-feedback paradigm following completion of the Implicit Association Test (Greenwald, McGhee, & Schwartz 1998). Finally, in Study 3, I independently manipulate "self-affirmation" and "collective bias" pre-feedback interventions to evaluate the relative success of each in reducing motivated reasoning processes and its effects.

General Overview of Analysis Across Studies

I examine the motivated rejection of evidence and allegations of implicit racial bias across three experiments in which I manipulate *general* beliefs about the existence of implicit bias (Study 1) and *specific* beliefs regarding one's own personal implicit bias (Study 2). This approach will allow me to assess the causal effects of allegations of implicit racial bias on motivated reasoning processes, as well as strategies to attenuate this defensiveness and its impact on behavior (Study 3). The experimental paradigm and materials were rigorously tested during a pilot study on an independent sample drawn from the same population (i.e., Amazon MTurkers) that will be used for the main studies. Although MTurk samples are not a representative, random sample of the American public, MTurk samples are older and more diverse than typical samples of university students, and more nationally representative than typical internet samples (Buhrmester, Kwang & Gosling 2011; Berinsky, Huber, & Lenz 2012; Mason & Suri 2012). By utilizing MTurk, I was able to obtain a large, non-random sample of White Americans with sufficient variability on demographic characteristics and, more importantly, the

constructs of interest (see Paolacci & Chandler, 2014, on the usefulness of MTurk for psychological research). All participants were recruited from Amazon MTurk to complete a study called, "Attitudes About People". This name was intended to increase the expectation that one's beliefs and attitudes about other people would be directly measured. For this reason, this name will be used for all of the pilots and the main experiments.

Analyses were conducted using ordinary least squares and binary logistic regression. Robust standard errors were used in all tests on coefficients. All variables were rescaled to run from 0-1 for easier comparison and estimation of effect sizes, unless otherwise noted, and experimental condition was dummy-coded. All models included age, gender, income, and education as covariates. A measure of trust in social science was also included as a covariate for analyses in the main studies to control for the influence of general skepticism or devaluation of the social sciences on estimated effects. Simple slope analyses for significant interactions between continuous variables were computed at one standard deviation above and below the mean of the moderator, following the procedures recommended by Aichen and West (1991), although most tests of moderation hypotheses involve examining the relationship of specifc constructs separately in each experimental condition. Finally, mediation analyses were conducted by following the steps for bootstrapping recommended by Preacher & Hayes (2004; also see Zhao, Lynch, & Chen, 2010).

Pilot 1a: General Belief in the Existence of Implicit Racial Bias

The experimental paradigm used in Study 1 was pilot tested on a sample drawn from the target population. Study 1 is designed to manipulate the general belief in the

existence of implicit racial bias in the population. The first pilot for Study 1 employed a single independent variable design (Implicit Bias Message vs. Control). The primary goal of the first pilot is to independently validate the experimental stimuli by assessing whether participants are able to comprehend information about implicit racial bias, that such information is perceived as believable, and that participants do not suspect an intention to deceive or mislead them about psychological research on implicit social cognition. In general, no differences between conditions were expected for the perceived believability, accuracy, or comprehensibility of the message prompt, or in the belief that the experimenters intend to deceive participants. Finally, pilot 1a also provides an opportunity to examine the impact of credible information about implicit racial bias on motivated reasoning processes.

Method

Participants

Participants were 269 White U.S. citizens recruited from Amazon MTurk (45% females, 55% males; mean age = 37.08, *SD* = 12.35). Most participants were modestly affluent (44%report a family income greater than 50K) and educated (54% have earned at least a Bachelor's degree). With this sample size, to detect mean level differences between the experimental and control group, I estimated that I had 50% power to detect a Cohen's *d* of 0.2, 99% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit White U.S. citizens and would compensate participants $0.50 for 20 minutes of their time.

After completing the consent procedure, participants first received the following instructions:

Below is part of an article that was published recently in one of the most prestigious psychological journals in the United States, *American Psychologist*. After many years of systematic research and debate on the issue, the view presented in the article is considered to be the established conclusion among social and behavioral scientists worldwide. The study's results have far-reaching consequences for a scientific understanding of human cognition and behavior, but it also has important implications for educators, businesses, lawyers, judges, journalists, politicians, and many other influential individuals and organizations. Please read the article summary carefully and answer the questions at the end of it.

Next, participants were randomly assigned to either the experimental or control condition. Information contained in both the experimental and control condition were designed to be similar in length and credibility. However, the experimental condition reviewed research on implicit racial bias and its consequences for racial minorities, whereas the control condition reviewed research on the health benefits associated with regular exercise:

Experimental Condition:

After decades of scientific research, a large body of evidence indicates that the overwhelming majority of Americans harbor unconscious forms of prejudice towards racial minorities. Social and behavioral scientists have concluded that unconscious racial attitudes can influence people's behavior without their awareness and without the ability to control its effects because it reflects the way we think about others. As a result, these unconscious beliefs lead to significant discrimination towards racial minorities in many important life domains, including education, employment, criminal justice, and more.

For example, one study demonstrated that White employers who scored higher on measures of unconscious racial bias evaluated White job applicants more positively than

Black job applicants, even though both job applications were identical. These findings have been replicated in many additional contexts and with many different participants.

Furthermore, psychologists also agree that unconscious racial bias exists even for people who consciously and deliberately reject all forms of racial bigotry- that is, people who consciously support racial equality often support negative stereotypes about racial minorities at the unconscious level because of how we think about others. In fact, the authors of a recently published study argued that "the vast majority of Americans are unconsciously prejudiced towards racial minorities, and this unconscious bias is the primary cause of racial inequality in our society. While most Americans consciously support racial equality, these same people can also be characterized as unconsciously prejudiced and discriminatory towards racial minorities." These experiments, and those of countless other researchers, have consistently found strong support for these claims. *(Word Count=236)*

Control Condition:

After decades of scientific research, a substantial body of evidence indicates that regular exercise and a healthy diet is essential to people's happiness and well-being. Social and behavioral scientists now agree that people who eat healthy and exercise regularly are much happier with their life. A healthy diet and active lifestyle helps people manage their stress, increase their energy levels, and improve their immune system.

For example, one study demonstrated that people who report eating healthy and exercising regularly were judged by their friends and family to be more happy and energetic. These findings have been replicated in many different studies and with many different people.

Furthermore, psychologists also agree that the benefits of a healthy diet and active lifestyle can be achieved immediately with only minor changes- that is, within a few short months, people who adopt a more healthy diet and active lifestyle report being less stressed, more energetic, and better able to cope with disease. In fact, the authors of a recently published study argued that "people who eat healthy and exercise regularly are significantly more happy and less stressed than people who do not eat healthy and exercise. While most Americans want to be happy, these same people commonly fail to make the necessary changes in their lives to achieve that goal." His experiments, and those of countless other researchers, have consistently found strong support for these claims. *(Word Count=232)*

Participants then completed a series of questions designed to measure their comprehension of the message prompt, the extent to which they found the information believable, and whether they suspected that the researchers had ulterior motives and

intentionally deceived them. Finally, participants completed the shortened motivated reasoning battery and demographic measures before being fully debriefed.

Measures

Reading Comprehension. Four items (3 true/false; 1 multiple choice) were used to evaluate participants' comprehension of the information contained in the message. These items differed, depending on condition assignment. Participants in the experimental message responded to such items as, "A large body of scientific evidence proves that even Americans who intend well and support racial equality may be unconsciously biased towards racial minorities". Participants in the control condition responded to such items as, "A large body of scientific evidence proves that Americans who eat healthy and exercise regularly are less stressed, more energetic, and better able to cope with illness". Correct responses to each item was coded as 1, whereas incorrect responses were coded as 0. I summed responses across all four items for both the experimental and control (M=.97, SD=.08) condition, which indicated high levels of comprehension.

Manipulation Checks. 5 items were used to validate the manipulation. On a 7-point scale ranging from 1 ("Strongly Disagree") to 7 ("Strongly Agree"), participants indicated whether 1) "The information presented to me was believable" and 2) "I believe the information presented to me was accurate". Participants also reported their confidence in each judgment. Perceived believability was indexed by the average of responses to the item and expressed confidence in that response (M=.77, SD=.17). Perceived accuracy was indexed the same way (M=.64, SD=.18). Participants also indicated if they thought "the researcher intentionally tried to deceiver or mislead you" ("No"=0, "Yes"=1; M=.18, SD=.39).

Shortened Motivated Reasoning Battery (SMR-B). 22 items were used to measure the motivated rejection of personal and general implicit racial bias, prejudice, and discrimination, concerns about implicit bias, willingness to participate in an anti-bias intervention, and perceptions of the credibility, validity, and extremism of social scientists. Responses across all 22 items were average to obtain a general index of motivate reasoning in this context (M=.55, SD=.90, Cronbach's alpha=.84), such that higher values represent greater motivated reasoning. However, 3 theoretically distinct subscales were also obtained, and allow for a more nuanced investigation of the specific beliefs relevant to these motivated reasoning processes. Accordingly, these subscales are the focus of my analyses. Each subscale is described below.

Self-Perceived Implicit Racial Bias. Participants reported a belief in their own implicit racial bias across 6-items. On a 7-point scale, participants responded to such items as, "How likely is it that your unconscious beliefs are unfavorable toward racial minorities?", "Do you believe that your unconscious racial attitudes influence your behavior towards racial minorities in an unfair way?", and "How worried are you that you are unconsciously prejudiced towards racial minorities?". Higher values represent increased belief in one's implicit racial bias (M=.50, SD=.25, Cronbach's alpha=.83).

Perceptions of Social Science. Participants reported their belief in the credibility, objectivity, integrity, and extremism of social science and scientists across 8-items. On a 7-point scale, participants responded to such items as, "To what extent are social scientists who study the psychology of unconscious racial bias motivated by a political or ideological agenda?", "How credible are social scientists who study the psychology of unconscious racial bias?", and "Social scientists think that all White

people are racists.". Higher values represent more favorable attitudes towards social science and scientists (M=.61, SD=.19, Cronbach's alpha=.84).

Belief in Racial Prejudice and Discrimination. Participants reported their belief in the existence of implicit racial bias, and racial prejudice, discrimination, and inequality in the general population across 7 items on a 7-point scale. Such items include 1) " How common is unconscious racial prejudice in America?", 2) "How likely is it that unconscious racial attitudes bias people's judgments and behavior towards racial minorities?", 3) "Do you think differences between racial groups can be explained by the effects of unconscious racial bias and prejudice?", 4) "How common is racial prejudice in America?", and 5) "Do you think that racial minorities are held back in society because of racial prejudice?". Higher values represent greater belief in the existence of implicit racial bias in the general population (M= .61, SD=.20, Cronbach's alpha=.90).

Demographics. Participants reported their age, gender, race, family income, and level of education.

Results

Reading Comprehension and Manipulation Check

Participants' total number of correct response (0 to 4) to the reading comprehension items was regressed on a dummy-coded variable for condition assignment (0=control, 1=experimental). Comprehension of the information contained in the message did not differ for participants in the experimental or control condition (p>.1).

However, participants' judgment of the believability of the information contained in the message differed across condition *(b* = -.09 (*95% CI* = *-.13*, -.05), *p* <.001, *d*=.5). Participants regarded the information contained in the control condition as more

believable than the information contained in the experimental condition. Furthermore, the effect of condition on the perceived accuracy of the information did not obtain significance (p>.1). However, participants were marginally more likely to report the use of deception in the experimental, compared to the control, condition *(Exp(b)* = 1.82 (*95% CI = .94, 3.52*), *p* =.08).

Main Effects of Experimental Manipulation on Motivated Reasoning

Results indicate that participants in the experimental (vs. control) condition were significantly more likely to believe that they harbored implicit racial bias *(b* = .15 (*95% CI = .09* .21), *p* < .001, *d*=.59), to hold *negative* attitudes towards social science (*b* = -.04 (*95% CI = -.09, .004*), *p* =.07 , *d*=.29), and marginally more likely to report a belief in the existence of racial prejudice and discrimination in the general population *(b* = .04 (*95% CI = -.004, .09*), *p* = .07 , *d*=.29). The size of these effects range from small to moderate in size. Given that all variables were recoded on a 0-1 interval, substantively these estimates indicate that participants in the experimental (vs. health control) conditions reported approximately a 15% increase in the belief in their own implicit racial bias, 4% decrease in favorable attitudes towards social science, and 4% increase in belief in racial prejudice. The estimated marginal means for each variable as a function of assignment to experimental or health control condition is provided in Figure 1.1.

Pilot 1a Discussion

Participants exposed to credible information about implicit racial bias and its consequences for racial inequality were more willing to believe that they harbored implicit racial bias, and were slightly more willing to recognize the existence of discrimination and prejudice in society, compared to participants exposed to information

about the health benefits of diet and exercise. However, participants exposed to credible information about implicit racial bias (vs. health and exercise) also expressed more negative attitudes towards social science and scientists.

More importantly, participants' comprehension of the information contained in both messages was very high (Full Sample Mean=3.84 out of 4), and did not differ across condition. Nevertheless, participants' regarded the information contained in the experimental (vs. control) condition as less believable and more deceptive. These differences confound the inferred impact of the experimental manipulation on relevant beliefs, which is especially problematic for analyses evaluating the role of individual differences on motivated reasoning processes in Study 1. For this reason, a second pilot of Study 1 was undertaken.

Pilot 1b: General Belief in the Existence of Implicit Racial Bias

Pilot 1b was designed to test a modified version of the materials included in the first pilot of Study 1 and to evaluate the validity of an alternative control group that is expected to be perceived as more similar to the experimental group. In particular, the stimuli used in pilot 1a were modified to increase the perceived believability of the experimental group, and adjust downward the perceived believability of the control group. Furthermore, an additional control group was included, which summarizes research on automaticity and the illusion of free will. The benefit of this message as a comparison to the experimental message is its focus on unconscious psychological functioning and its effect on social judgment and behavior, without directly implicating intergroup attitudes, discrimination, or racial inequality.

The second pilot for Study 1 employed a single independent variable design (Implicit Bias Message vs. Health vs. Automaticity). The primary goal of the second pilot is to independently validate the experimental stimuli by assessing whether participants are able to comprehend information about implicit racial bias, that such information is perceived as believable, and that participants' do not suspect an intention to deceive or mislead them about psychological research on implicit social cognition. In general, no differences between conditions were expected for the perceived believability, accuracy, or comprehensibility of the message prompt, or in the belief that the experimenters intend to deceive participants. Finally, pilot 1b also provides an opportunity to examine the impact of credible information about implicit racial bias on motivated reasoning processes, and to replicate the results obtained in pilot 1a.

Method

Participants

Participants were 387 White U.S. citizens recruited from Amazon MTurk (56% females, 46% males; mean age = 37.11, *SD* = 13.22). Most participants were modestly affluent (41% report a family income greater than 50K) and educated (52% have at least earned a Bachelor's degree). With this sample size, to detect mean level differences between the experimental and a single control group, I estimated that I had 48% power to detect a Cohen's *d* of 0.2, 99% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

The procedure of pilot 1b was identical to the procedure for pilot 1a. However, participants assigned to the new control group read the following information:

After decades of rigorous scientific research, a large body of systematic evidence indicates that most people are far less aware of why they think and act the way that they do than they realize. Social and behavioral scientists have concluded that people are often mistaken about or unaware of what influences their beliefs and what influences their behavior. As a result, accurate awareness and control over thought and behavior may be an illusion.

For example, one large study demonstrated that exposing hundreds of people to subliminal images of "rats", "spiders", and "insects" led them to dislike their peers and to judge them more harshly- even though participants had no conscious awareness of being exposed to these images. Another study found that when hundreds of people were exposed to images of "nursing homes" and the "elderly", they walked away from the laboratory more slowly and with worse posture than those not exposed to such images. These findings have been replicated in dozens of different studies and contexts, and with thousands of different participants throughout the world.

In general, most psychologists agree that much of human thought and behavior is directed by automatic factors that are inaccessible to conscious awareness. In fact, the authors of a recently published study in a peer-reviewed journal conclude that "free will and conscious control appear to be unrelated to what people think and how people act. While most people believe that they have direct control over their thoughts and their behavior, these same people are completely unaware of how other factors more directly influence their beliefs and actions." These experiments, and those of other independent researchers, have found strong and consistent support for these claims. *(Word Count=279)*

Measures

The same measures used in pilot 1a were also used in pilot 1b. See Table 1.1 for the mean, SD, and Cronbach's alpha, when appropriate, for each measure used in analyses.

Results

Two dummy-coded variables were created to represent condition assignment, with the implicit bias condition as the reference group. For analyses comparing differences between the two control groups, the health control condition serves as the reference group. For all of the models used in analyses reported below, both dummy-coded variables are included along with covariates.

Reading Comprehension and Manipulation Check

Participants' total number of correct response to reading comprehension was regressed on two dummy-coded variables for condition assignment. Comprehension of the information contained in the message did not differ for participants in the experimental condition compared to automaticity control condition (p>.1). However, reading comprehension was slightly higher in the health control compared to the experimental condition *(b* = .11 (*95% CI = .00*, .23), *p* = .059, *d*=.23). Furthermore, participants' judgment of the believability, accuracy, and deceptiveness of the information contained in the message did not differ between the experimental condition and the automaticity control condition (p>.1), providing evidence to support the validity of using these stimuli for Study 1.

However, consistent with the results obtained in pilot 1a, participants judgment of the believability of the information contained in the message differed between the experimental and health control condition *(b* = .13 (*95% CI =.09, .17*), *p* <.001, *d*=.74). Participants regarded the information contained in the health control condition as more believable than the information contained in the experimental condition. While the effect of condition on the perceived accuracy of the information did not obtain significance (p>.1), participants were, once again, significantly more likely to report the use of deception in the experimental, compared to the health control, condition *(Exp(b)* = .23 (*95% CI = .11, .47*), *p* <.001).

Main Effects of Experimental Manipulation on Motivated Reasoning

Consistent with pilot 1a, results indicate that participants in the experimental, compared to the healthy control, condition were more likely to believe that they harbored implicit racial bias *(b* = -.07 (*95% CI = -.13*, -.51), *p* < .05, *d*=.26) and to hold *negative*

attitudes towards social science (*b* = .06 (*95% CI = .01, .11*), *p* <.05 , *d*=.29). However, unlike pilot1a, no significant differences emerged for belief in the existence of racial prejudice and discrimination in the general population (p>.1). Given that all variables were recoded on a 0-1 interval, substantively these estimates indicate that participants in the experimental (vs. health control) conditions reported approximately a 7% increase in the belief in their own implicit racial bias and a 6% decrease in favorable attitudes towards social science. The estimated marginal means for self-perceived racial bias, perceptions of social science, and belief in prejudice as a function of assignment to experimental or health control condition is provided in Figure 1.2.

Next, I evaluated differences between the experimental and automaticity control condition. Results indicate that participants in the experimental, compared to the automaticity control, condition were more likely to hold *negative* attitudes towards social science (*b* = .01 (*95% CI = .05, .15*), *p* <.001 , *d*=.48). No other differences between these two conditions obtained significance (p>.1). Furthermore, no significant differences between the health and automaticity control conditions were observed (p>.1).

Pilot 1b Discussion

Consistent with pilot 1a, participants regarded the information contained in the health control condition as more believable and less deceptive than the experimental condition. These differences disqualify the health condition from inclusion in Study 1. However, despite the perceived differences in the believability and deceptiveness of the information contained in both conditions, participants in the experimental (vs. health control) condition were significantly more likely to believe they harbored implicit bias and to express negative attitudes towards social science. This finding was substantively

meaningful, as it accounted for approximately a 6-7% difference between groups. Thus, self-perceived implicit bias appears to be a malleable attitudinal orientation, capable of changing in relation to, and in the direction of, credible scientific information. Conversely, pilot 1b failed to replicate the effect of the experimental (vs. health control condition) on belief in the existence of racial prejudice and discrimination in the general population, which was observed in pilot 1a. Efforts to communicate to the public scientific information about implicit racial bias, while capable of increasing self-awareness of implicit bias, may nevertheless risk backfire effects against social scientists, and may also fail to reliably promote inferences about the existence of racial prejudice and discrimination in society.

In contrast, participants exposed to credible information about automatic psychological functioning were just as able to perceive their own implicit racial bias, and no more inclined to report a belief in the existence of racial prejudice and discrimination in the general population, as participants in the experimental group. However, attitudes towards social science were more favorable in the automatic control (vs. experimental) condition. Nonetheless, I failed to obtain any clear evidence to suggest that participants in the automaticity control condition were more able to perceive their own implicit racial bias or recognize the existence of the racial prejudice and discrimination in society, compared to the health control condition.

More importantly for the purposes of pilot 1b, no significant differences in the perceived believability, accuracy, or deceptiveness of the information contained in the message were observed between participants assigned to the experimental condition and automaticity condition. Comprehension of the information contained in both conditions

was also high and did not differ. Accordingly, the stimuli materials contained in these conditions will be used for Study 1. Study 1 allowed me to examine the role of individual differences and its interaction with experimental condition, and to investigate downstream effects on prejudice-regulation, stereotyping, and public policy attitudes.

Study 1: General Belief in the Existence of Implicit Racial Bias

Study 1 employed a single independent variable design (Implicit Bias Message vs. Automaticity Control), and is designed to manipulate the belief in the existence of implicit racial bias in the general population. Half of the participants were randomly assigned to read an excerpt from a scientific article that summarizes research on implicit racial bias, whereas the other half of participants read an article that summarizes research on automatic psychological functioning and social cognition. This Study used a 2-wave panel design. At Time 1 (T1), participants completed a survey designed to measure their explicit racial attitudes, egalitarian motivations, sociopolitical orientations, and trust in social science. At Time 2 (T2), participants were randomly assigned to experimental condition, before completing a survey that measured motivated reasoning processes (SMR-B), attitudes towards anti-bias interventions, motivation to control prejudice reactions, endorsement of racial stereotypes, and attitudes toward public policy intended to help racial minorities.

The purpose of Study 1 was to evaluate the psychological factors associated with the motivated rejection of implicit racial bias, and their consequences for prejudice regulation, stereotyping, and public policy attitudes. In particular, it was expected that people characterized by explicitly negative racial attitudes, resentment towards racial progress, authoritarianism, system justification, or social dominance orientation would be

defensively motivated to 1) reject evidence of aggregate-level racial prejudice and discrimination, 2) judge social science and scientists unfavorably, and 3) to be less able to perceive their own implicit racial biases (Hypothesis 1). These individuals are also expected to be 1) less supportive of anti-bias interventions, 2) less motivated to control prejudice reactions, 3) more inclined to endorse racial stereotypes, and 4) opposed to racially liberal public policies. Critically, the hypothesized effects of explicit racial attitudes and sociopolitical orientations on motivated reasoning and stereotyping are expected to be more pronounced among participants exposed to credible information about implicit racial bias, compared to participants in the automaticity control condition (Hypothesis 2).

Furthermore, intrinsically motivated egalitarians were expected to be more willing- and extrinsic egalitarians less willing- to 1) accept evidence of aggregate-level racial prejudice and discrimination, 2) judge social science and scientists favorably, and 3) perceive their own implicit racial biases. I expect this pattern of results to be most pronounced among participants in the implicit bias (vs. automaticity control) condition (Hypothesis 3).

Because T1 independent variables are expected to 1) predict motivated reasoning and stereotyping, and 2) motivated reasoning is also expected to predict attitudes toward anti-bias interventions (Hypothesis 10), general motivation to control prejudice reactions (Hypothesis 11), and stereotyping (Hypothesis 7), I also tested for mediation. It is expected that sociopolitical orientations, explicit racial attitudes, and extrinsic egalitarian motivations will increase negative attitudes towards social scientists and decrease self-perceived bias and belief in racial prejudice. These constructs are expected, in turn, to

increase racial stereotyping. Finally, I also expected stereotyping to predict attitudes toward public policies intended to help minorities (Hypothesis 8 and 9).

Method

Participants

Participants were 360 White U.S. citizens recruited from Amazon MTurk (58% females, 42% males; mean age = 33.92, *SD* = 11.33). Most participants were modestly affluent (37% report a family income greater than 50K) and educated (48% have at least earned a Bachelor's degree). Of these individuals, 278 were retained at T2 (77%). With this sample size (n=277), to detect mean level differences between the experimental and control group, which was assessed at T2, I estimated that I had 46% power to detect a Cohen's *d* of 0.2, 99% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8. 5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit white U.S. citizens for a two-part study, and would compensate participants $1.25 total for their time (T1=$0.25, T2=$1.00).

Pre-manipulation measures were assessed at T1. Approximately one week later, participants were contacted to complete the second part of the study. At T2, participants were randomly assigned to either the experimental or health control condition (described in pilot 1b). Next, participants completed the post-manipulation battery and demographic measures before being fully debriefed.

Measures

Sociopolitical Orientations

Social Dominance Orientation (SDO). The social dominance orientation scale (version 6) (Sidanis & Pratto, 2001) consists of 16 items to which the participants are asked to state their degree of agreement on a 7-point scale to such items as, "Some groups of people are simply inferior to other groups." Higher values represent higher levels of social dominance orientation (Cronbach's alpha=.93).

System Justification Participants reported their belief in the legitimacy of the status quo (Kay & Jost, 2003) across 8 items. On a 9-point scale (1="Strongly Disagree" to 9="Strong Agree"), participants responded to such items as, "In general, you find society to be fair". Higher values representation higher levels of system justification (Cronbach's alpha=.82).

Explicit Racial Attitudes

Racial Resentment. The racial resentment scale measures participants' explicit belief that blacks are unable or unwilling to work hard enough to overcome obstacles to success and are therefore undeserving of assistance or special favors (Kinder & Sanders, 1996). Participants responded to 4-items on a 5-point scale (1=disagree strongly, to 5=agree strongly), such as "It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites". Higher values represent higher levels of racial resentment (Cronbach's alpha=.76).

Attitudes Toward Blacks Scale. Participants reported their explicit attitudes towards blacks (Brigham, 1993) across 20 items on a 7-point scale (1=strongly agree, 7=strongly disagree). Such items include "Black and white people are inherently equal" and "It would not bother me if my new roommate was black". Higher values were coded to represent relatively more negative attitudes towards blacks (Cronbach's alpha=.91).

Egalitarian Motivations

Intrinsic and Extrinsic Egalitarian Motivations. Participants reported the extent to which they are internally or externally motivate to control their prejudice (Plant & Devine, 1998). External motivations were measured across 5 items on 9-point scale (1=strongly disagree to strongly agree), including such items as " I try to act nonprejudiced toward Black people because of pressure from others". Higher values on this scale correspond with increased extrinsic egalitarian motivations (Cronbach's alpha=.90). Internal motivations were measured across 5 items on 9-point scale (1=strongly disagree to strongly agree), including such items as " I am personally motivated by my beliefs to be unprejudiced toward Black people". Higher values on this scale correspond with increased intrinsic egalitarian motivations (Cronbach's alpha=.91). A difference score was computed by subtracting scores on extrinsic egalitarian motivations from scores on intrinsic egalitarians, such that higher values represent increased intrinsic vs. extrinsic egalitarian motivations.

Shortened Motivated Reasoning Battery (SMR-B). The same 22 items were used to measure the motivated rejection of personal and general implicit racial bias, prejudice, and discrimination, concerns about implicit bias, willingness to participate in an anti-bias intervention, and perceptions of the credibility, validity, and extremism of social scientists. Responses across all 22 items were averaged to obtain a general index of motivated reasoning in this context (M=.59, SD=.13, Cronbach's alpha=.84), such that lower values represent greater motivated reasoning. As before, 3 theoretically distinct subscales were also obtained, and allow for a more nuanced investigation of the specific beliefs relevant to these motivated reasoning processes. Accordingly, these subscales are

the focus of my analyses. Each subscale is described below.

Self-Perceived Implicit Racial Bias. Participants reported a belief in their own implicit racial bias across 6-items. On a 7-point scale, participants responded to such items as, "How likely is it that your unconscious beliefs are unfavorable toward racial minorities?", "Do you believe that your unconscious racial attitudes influence your behavior towards racial minorities in an unfair way?", and "How worried are you that you are unconsciously prejudiced towards racial minorities?". Higher values represent increased belief in one's implicit racial bias (Cronbach's alpha=.84).

Perceptions of Social Science. Participants reported their belief in the credibility, objectivity, integrity, and extremism of social science and scientists across 8-items. On a 7-point scale, participants responded to such items as, "To what extent are social scientists who study the psychology of unconscious racial bias motivated by a political or ideological agenda?", "How credible are social scientists who study the psychology of unconscious racial bias?", and "Social scientists think that all White people are racists.". Higher values represent more favorable attitudes towards social science and scientists (Cronbach's alpha=.84).

Belief in Racial Prejudice and Discrimination. Participants reported their belief in the existence of implicit racial bias, and racial prejudice, discrimination, and inequality in the general population across 7 items on a 7-point scale. Such items include 1) " How common is unconscious racial prejudice in America?", 2) "How likely is it that unconscious racial attitudes biases people's judgments and behavior towards racial minorities?", 3) "Do you think differences between racial groups can be explained by the effects of unconscious racial bias and prejudice?", 4) "How common is racial prejudice in

America?", and 5) "Do you think that racial minorities are held back in society because of racial prejudice?". Higher values represent greater belief in the existence of implicit racial bias in the general population (Cronbach's alpha=.89).

Prejudice Regulation

*Motivation to Control Prejudice Responses (*MCPR). Participants reported their MCPR (Dunton & Fazio, 1997) across 12 items on a 7-point scale (1=strongly agree, 7=strongly disagree). Scores were recoded such that higher values represent higher levels of MCPR (Cronbach's alpha=.76).

Attitudes Towards Anti-Bias Interventions. Participants reported their attitudes towards interventions to reduce implicit bias in organizational contexts, law enforcement, and society generally on a 7-point scale (1=not at all important/valuable, 7= Extremely important/valuable). Responses across 6 items were obtained, and included such items as, "In your opinion, how important is it that the law requires employers to undergo training to reduce their unconscious racial bias?", "In your opinion, how important is it that law enforcement personnel undergo training to reduce their unconscious racial bias?", and "In your opinion, how valuable are anti-bias programs in eliminating racial discrimination in the society?". Higher values represent more positive attitudes towards anti-bias interventions (Cronbach's alpha=.95).

Stereotype Endorsement. Participants were asked to attribute racial disparities in the criminal justice system, employment, and socioeconomic status to either personal or situational factors. Responses were provided to 2-items for each of the 3 stereotype domain (6-items total) on a 7 point scale ranging from situational attributions to personal attributions. Such items included, "In your view, why are Blacks more likely to live in

poverty than Whites?", "In your view, why are Blacks less likely to be hired and promoted than Whites?, and "In your view, why are Blacks more often arrested and sent to prison than Whites?". Responses were averaged across all 6-items, and coded such that higher values represent greater endorsement of racial stereotypes (Cronbach's alpha=.95).

Public Policy Attitudes. Participants reported their attitudes about criminal justice policies, affirmative action, and social welfare (Peffley & Hurwitz, 2010; Gilens, 1997) across 9-items on a 7-point scale, ranging from support or opposition to public policies intended to help racial minorities. Such items included, "When people can't support themselves, the government should help by giving them enough money to meet their needs.", "Do you strongly approve, somewhat approve, somewhat disapprove, or strongly disapprove of racial profiling?", and "Do you think that the government should pass laws to protect racial minorities from job discrimination?". Higher values represent greater support for public policy intended to help racial minorities (Cronbach's alpha=.81).

Skepticism About Social Science. Participants reported their general attitudes towards social science across 4-items, adapted from McCright, Dentzman, Charters, and Dietz (2013). Participants responded on a 5-point scale (1=completely distrust, 5=completely trust) to such items as, "How much do you distrust or trust social scientists to create knowledge that is unbiased and accurate?". Higher values represent more favorable opinions of social science (Cronbach's alpha=.89).

Demographics. Participants reported their age, gender, race, family income, and level of education.

See Table 1.2 for means, SD, and correlations between all measures used in

analyses for Study 1.

Results

Two dummy-coded variables were created to represent condition assignment, with the automaticity condition as the reference group.

Effects of Experimental Manipulation on Motivated Reasoning

No significant main effects of experimental condition was observed on self-perceived implicit racial bias, attitudes towards social science, or beliefs in the existence of racial prejudice and discrimination ($p<.1$). The failure to detect main effect differences between groups is not surprising given that the stimuli used in this experiment were chosen to be highly similar in content, comprehensibility, believability, accuracy, and deceptiveness.

Effects of Individual Differences on Motivated Reasoning

To evaluate the psychological predictors of the motivated rejection of implicit racial bias, each subscale of SMR-B was regressed separately on each indicator of explicit racial attitudes, sociopolitical orientations, and egalitarian motivations. Figure 1.3 provide a coefficients plot that graphically represent the unstandardized coefficients and 95% CI for each independent variable obtained from models estimating self-perceived implicit racial bias, perceptions of social scientists, and belief in prejudice and discrimination, respectively. Below I describe the results of these analyses.

I begin by first examining the impact of explicit racial attitudes. Results indicate that racial resentment was associated with negative attitudes toward social science (b = -.18 (*95% CI =-.26, -.10*), p <.001) and a decreased belief in racial prejudice and discrimination (b = -.28 (*95% CI = -.35, -.21*), p <.001), but was unrelated to self-

perceived racial bias (p>.15). Similarly, explicitly negative racial attitudes was associated with negative attitudes toward social science (*b* = -.27 (*95% CI -.42, -.13*), *p* <.001), a decreased belief in racial prejudice and discrimination (*b* = -.37 (*95% CI = -.55, -.19*), *p* <.001), and, unlike racial resentment, increase in self-perceived racial bias (*b* = .25 (*95% CI =.06, .45*), *p* <.001).

Next, I estimated the effect of sociopolitical orientations on motivated reasoning processes. These results indicate that system justification was associated with negative attitudes toward social science (*b* = -.14 (*95% CI =-.26, -.02*), *p* <.05) and a decreased belief in racial prejudice and discrimination (*b* = -.36 (*95% CI = -.46, -.25*), *p* <.001), but was unrelated to self-perceived racial bias (p>.15). In contrast, social dominance orientation predicted negative attitudes toward social science (*b* = -.30 (*95% CI = -.42, -.18*), *p* <.001), a decreased belief in racial prejudice and discrimination (*b* = -.30 (*95% CI = -.43, -.16*), *p* <.001), and, unlike system justification, increase in self-perceived racial bias (*b* = .17 (*95% CI =.01, .33*), *p* <.05).

A parallel set of analyses was undertaken to examine the effects of egalitarian motivations. Results indicate that intrinsic (vs. extrinsic) egalitarian motivations predicted decreased self-perceived implicit racial bias (*b* = -.36 (*95% CI = -.52, -.19*), *p* <.001), but more favorable perceptions of social scientists (*b* = .25 (*95% CI = .14, .36*), *p* <.001), and increased belief in discrimination and prejudice (*b* = .19 (*95% CI = .08, .30*), *p* =.001).

Finally, to determine whether individual differences were activated by the experimental condition to influence motivated reasoning processes, interaction terms were constructed by each T1 measure and the dummy-coded variable for experimental

condition, and subjected to OLS. This analysis revealed that the effect of explicit racial attitudes, sociopolitical orientation, and egalitarian motivations on motivated reasoning processes did not vary across experimental condition (p>.1).

Anti-Bias Interventions, Prejudice-regulation, Stereotyping, and Public Policy Attitudes

The next set of analyses examines the relationship between 1) each subscale of the SMR-B, explicit racial attitudes, sociopolitical orientations, and egalitarian motivations, and 2) attitudes towards anti-bias interventions, prejudice-regulation, stereotyping, and public policy attitudes. Each independent variable was regressed separately on each dependent variable using ordinary least squares regression. Figure 1.4 and 1.5 provides a coefficients plot that graphically represent the unstandardized coefficients and 95% CI for each independent variable obtained from each model.

These results clearly indicate that perceptions of social scientists, belief in prejudice and discrimination, sociopolitical orientations, and egalitarian motives are each significant predictors of anti-bias interventions, stereotyping, and public policy attitudes in ways consistent with my theoretical expectations. In particular, self-perceived bias (*b* = .16 (*95% CI =.03, .29*), *p* =.013), favorable attitudes towards social sciences (*b* = .28 (*95% CI =.08, .48*), *p* =.007), belief in discrimination and prejudice (*b* = .88 (*95% CI =.69, 1.06*), *p* <.001), and intrinsic (vs. extrinsic) egalitarianism (*b* = .46 (*95% CI =.31, .61*), *p* <.001) was associated with more positive attitudes towards anti-bias interventions; racial resentment (*b* = -.33 (*95% CI = -.45, -.21*), *p* <.001), negative black attitudes (*b* = -.61 (*95% CI =-.85, -.37*), *p* <.001), SDO (*b* = -.48 (*95% CI =-.65, -.30*), *p* <.001), and system justification (*b* = -.33 (*95% CI =-.51, -.15*), *p* <.001) were associated with less

favorable attitudes towards anti-bias interventions. Only increased perceptions of social science (*b* = .10 (*95% CI =-.02, .21*), *p* =.096), belief in prejudice and discrimination (*b* = .14 (*95% CI =.03, .25*), *p* =.014), decreased racial resentment (*b* = -.10 (*95% CI =-.17, -.03*), *p* =.004), and positive attitudes towards blacks (*b* = -.11 (*95% CI =-.22, .01*), *p* =.068) was associated with the motivation to control prejudiced reaction.

A similar pattern of results emerged for racial stereotyping and public policy attitudes. In particular, favorable perceptions of social scientists (*b* = -.51 (*95% CI = -.69, -.32*), *p* < .001), belief in prejudice and discrimination (*b* = -.79 (*95% CI =-.97, -.61*), *p* < .001), and intrinsic (vs. extrinsic egalitarianism ((*b* = -.61 (*95% CI = -.76, -.46*), *p* <.001) was associated with less racial stereotyping; racial resentment (*b* = .66 (*95% CI =.58, .74*), *p* <.001), negative attitudes towards blacks (*b* = 1.06 (*95% CI =.91, 1.22*), *p* < .001), SDO (*b* = .86 (*95% CI =.73, .98*), *p* <.001), and system justification was associated with increased stereotyping (*b* = .62 (*95% CI =.44, .80*), *p* <.001). Similarly, favorable perceptions of social scientists (*b* = -.26 (*95% CI =-.40, -.12*), *p* <.001), belief in prejudice and discrimination (*b* = -.58 (*95% CI = -.73, -.44*), *p* <.001), and intrinsic (vs. extrinsic egalitarianism ((*b* = -.42 (*95% CI =-.52, -.32*), *p* <.001) was associated with opposition to public policies intended to help racial minorities; racial resentment (*b* = .46 (*95% CI =.39, .52*), *p* <.001), negative attitudes towards blacks (*b* = .62 (*95% CI =.44, .79*), *p* <.001), SDO (*b* = .49 (*95% CI =.37, .61*), *p* <.001), and system justification (*b* = .41 (*95% CI =.30, .53*), *p* <.001) was associated with opposition to public policies intended to help racial minorities.

To test the hypothesis that sociopolitical orientations and explicit racial attitudes were activated by the experimental condition to influence stereotyping, interaction terms

were constructed by each T1 measure and the dummy-coded variable for experimental condition, and subjected to OLS. This analysis revealed that the effect of racial resentment on stereotype endorsement did not vary across experimental condition (p>.1).

Consistent with hypothesis 2, however, the interaction between experimental condition and social dominance orientation (*b* = -.26 (*95% CI =-.48, -.05*), *p* = .018), system justification (*b* = -.37 (*95% CI =-.73, -.01*), *p* = .046), and attitudes towards blacks (*b* = -.25 (*95% CI =-.52, -.02*), *p* = .073) obtained significance on stereotype endorsement. To determine if the effect of social dominance orientation, system justification, and attitudes towards blacks on stereotyping is stronger in the experimental (vs. control) condition, regression analyses were conducted separately for participants exposed to information about implicit racial bias or automaticity in psychological functioning. Inspection of the coefficients obtained in this analysis indeed indicate that the effect of social dominance orientation on stereotyping was larger in the experimental condition (*b* = .01 (*95% CI =.01, .01*), *p* <.001) compared to the control condition (*b* = .01 (*95% CI =.01, .01*), *p* <.001). The same results were observed for the effect of system justification ((*b* = .01 (*95% CI =.01, .02*), *p* <.001) vs. (*b* = .44 (*95% CI =.002, .01*), *p* <.001)) and attitudes towards blacks ((*b* = .01 (*95% CI =.01, .01*), *p* <.001) vs. (*b* = .01 (*95% CI =.01, .01*), *p* <.001)). These relationships are represented graphically in Figure 1.6.

Mediation Analyses for Stereotype Endorsement and Prejudice-Regulation

Because perceptions of social scientists and belief in racial prejudice and discrimination each predict stereotype endorsement, and are predicted by a broad range of variables that were found to also associate with stereotyping, these constructs represent

good candidates for mediation analysis. In order to examine whether each dimension of motivated reasoning mediates the relationship between 1) explicit racial attitudes, sociopolitical orientations, egalitarian motivations, and stereotype threat, and 2) stereotype endorsement, I ran a bootstrap analysis that tested for the indirect effect of each predictor on the dependent variable, retaining all covariates. Stereotype endorsement was chosen as the dependent variable because it is a common basis upon which White Americans justify racial inequality (e.g., Peffley & Hurwitz, 2010), and, in the present work, was expected and found to be an important predictor of attitudes toward public policy intended to help racial minorities, consistent with other research (e.g., Gilens, 1999). Evidence that these dimensions of the motivated rejection of implicit racial bias mediates the relationship between individual differences and stereotype endorsement would further illustrate its important role in organizing perceptions, and justifications, of implicit racial bias and discrimination. Such a pattern of evidence would also underscore the utility of anti-bias interventions in targeting these attitudes. Figure A provides a conceptual model of the expected pattern of mediation, in which beliefs about prejudice and discrimination and attitudes about social science each serve as a mediator for the relationship between individual differences and prejudice-regulation, stereotyping, and public policy attitudes.

The results of this analysis indicate that belief in racial prejudice and, to a lesser extent, perceptions of social scientists were each a critical mediator for the effects of individual differences on stereotype endorsement. Specifically, with belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects on stereotype endorsement for system justification *(b = .22, SE = .05, (95% CI = .12, .32),*

p<.001*),* social dominance orientation *(b = .19, SE = .04, (95% CI = .12, .29), p*<.001*),* racial resentment *(b = .14, SE = .03, (95% CI =.08, .20), p*<.001*),* attitudes toward blacks *(b = .23, SE = .05, (95% CI =.14, .34), p<.001),* and intrinsic (vs. extrinsic) egalitarian motivation *(b = -.13, SE = .002, (95% CI =-.20, -.05), p*<.001*).* Similarly, with perceptions of social scientists submitted as a mediator, I observe significant indirect effects on stereotype endorsement for social dominance orientation *(b = .10, SE = .04, (95% CI = .03, .18), p*<.01*),* racial resentment *(b = .07, SE = .02, (95% CI =.03, .12), p*<.01*),* attitudes toward blacks *(b = .07, SE = .03, (95% CI =.02, .14), p<.05),* system justification *(b = .06, SE = .03, (95% CI =.004, .11), p<.05),* and intrinsic (vs. extrinsic) egalitarian motivation *(b = -.08, SE = .03, (95% CI =-.15, -.03), p*<.01*).*

Because perceptions of social science and belief in the existence of racial prejudice were also important predictors of both indicators of prejudice-regulation, I also examine the extent to which these constructs mediate the relationship between sociopolitical orientations, racial attitudes, or egalitarian motivations on attitudes towards anti-bias interventions and general motivation to control prejudice reactions using the same procedure for mediation analyses described above.

The results of this analysis indicate that belief in racial prejudice, but not perceptions of social scientists, was a critical mediator for the effects of individual differences on intervention attitudes and general motivations to control prejudice reactions. Specifically, with belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects on attitudes towards anti-bias interventions for system justification *(b = -.31, SE = .06, (95% CI = -.42, -.20), p*<.001*),* social dominance orientation *(b = -.23, SE = .06, (95% CI = -.35, -.11), p*<.001*),* attitudes

towards blacks (b=-.29, SE= *.08* (95% CI =-.44, -.14), p<.001), racial resentment *(b = -.23, SE = .04, (95% CI =-30, -.16), p*<.001*),* and intrinsic (vs. extrinsic) egalitarian motivation *(b = .15, SE = .05, (95% CI =.06, .25), p*<.001*)*. However, with perceptions of social scientists submitted as a mediator, I did not observe any significant indirect effects on intervention attitudes.

For motivations to control prejudice reactions, perceptions of social scientists appears to be a less consistent mediator for the effect of individual differences on the general motivation to control prejudice reactions than for other mediation analyses. With belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects for system justification *(b = -.06, SE = .02, (95% CI =-.10, -.01), p=.011),* social dominance orientation *(b = -.04, SE = .02, (95% CI = -.08, -.002), p*=.036*),* attitudes towards blacks (b=-.04, SE= *.02* (95% CI =-.09, .003), p=.07), intrinsic (vs. extrinsic) egalitarian motivation *(b = .03, SE = .01, (95% CI =.004, .06), p*=.004*),* but not racial resentment (p>.15) or attitudes towards blacks (p>.1). However, with perceptions of social scientists submitted as a mediator, I did not observe any significant indirect effects on general motivations to control prejudiced reactions.

As expected, the indirect effects of individual differences through self-perceived implicit racial bias were not significant (p>.15). For this reason, I do not treat this variable as a mediator for parallel mediation analyses in Study 2.

Study 1 Discussion

Study 1 was designed to manipulate the belief in the existence of racial prejudice and discrimination in the general population. Using experimental stimuli designed to be similar in believability, accuracy, and deceptiveness, no reliable evidence was observed

in Study 1 to suggest that these beliefs, or the other dimensions of the motivated rejection of implicit bias, among White Americans differed as a function of exposure to credible information about implicit racial bias (vs. information on the automaticity of psychological functioning).

These results stand in contrast to what was observed for pilot 1a and 1b. Indeed, comparison of the main effects of exposure to credible information about implicit bias versus information about the health benefits of exercise and dieting, two highly disparate areas of psychological research, yielded a different pattern of results. Results across two independent samples recruited for pilot1a and pilot 1b both indicate that self-perceived implicit racial biases are malleable and capable of changing in relation to credible information, although the presentation of information on the existence of implicit racial bias also increased unfavorable attitudes towards social scientists and science. These effects emerged despite the increased tendency of participants in both samples to regard information in the health control condition as more believable and less deceptive than information contained in the implicit bias condition. Despite their skepticism of the information contained in the experimental condition and their relatively unfavorable attitudes towards social science that emerged, participants' self-perceived implicit racial bias still acquiesced in the direction of the scientific information that was presented.

This pattern of results raises the possibility that discussing automaticity in psychological functioning and social cognition, without directly implicating race or discrimination, may be a relatively more effective approach to communicating to White Americans the evidence for implicit racial bias and its consequences for discrimination and inequality. For example, the only differences between the automaticity control and

implicit bias condition observed in pilot 1b was an increased tendency in the former to hold more favorable attitudes towards social scientists, although this effect was not replicated in Study 1. Some caution here is clearly warranted, however, given that no differences were observed between the automaticity condition and health condition in pilot 1b or implicit bias condition in Study 1. These findings suggest that anti-bias interventions that make exclusive use of research on automaticity in psychological functioning and social cognition as a basis for educating the public on the existence and consequences of implicit racial bias may be inadequate. Additional work is needed to better explore this possibility.

Nevertheless, that self-perceived implicit racial bias is not necessarily a fixed attitudinal orientation is an important finding that substantiates the utility of anti-bias interventions that focus on increasing self-awareness of the propensity for prejudice-related discrepancies. However, across all 3 samples, beliefs about prejudice and discrimination in society were relatively unchanged in the face of credible information about implicit racial bias and its linkages to discrimination in the general population. This finding might suggest that general beliefs about prejudice and discrimination are more stable than self-perceived implicit bias, and may therefore reflect an orientation that render many White Americans motivated to justify, rationalize, or altogether ignore the reality of racial discrimination in society. Indeed, belief in prejudice and discrimination, but not self-perceived bias, mediated the relationship between individual differences and prejudice-regulation.

Accordingly, Study 1 was also designed to examine individual differences in the motivated rejection of implicit racial bias, and the consequences of these beliefs for

prejudice-regulation, stereotyping, and public policy attitudes. Specifically, I predicted that system justification, social dominance orientation, racial resentment, attitudes towards blacks, and extrinsic (vs. intrinsic) egalitarian motivations would each be associated with 1) decreased self-perceived implicit bias, belief in the existence of racial discrimination, favorable attitudes towards social science, motivation to control prejudice, and valuation of anti-bias interventions, and 2) increased stereotyping and opposition to public policies intended to help racial minorities. In contrast, I expected that people characterized by intrinsic egalitarian motivations would be 1) more likely to perceive their own implicit bias, believe in racial prejudice and discrimination, hold favorable attitudes towards social science, be motivated to control prejudiced responding, and value anti-bias interventions, and 2) less likely to endorse racial stereotypes or oppose public policies intended to help racial minorities. Furthermore, I expected that self-perceived implicit bias, belief in the existence of racial prejudice, and favorable attitudes towards social scientists would each predict increased motivations to control prejudice reactions, valuation of anti-bias interventions, opposition to racial stereotypes, and support for public policy intended to help racial minorities. Results largely support these predictions. Thus, consistent with my hypotheses, sociopolitical orientations, explicit racial attitudes, and egalitarian motivations are clearly important determinants of the motivated rejection or acceptance of implicit racial bias, which predicted prejudice-regulation, racial stereotyping, and public policy attitudes in theoretically expected ways.

The effect of explicit racial attitudes, sociopolitical orientations, and egalitarian motivations on motivated reasoning processes did not vary across experimental condition, suggesting that these individual differences may promote relatively stable

attitudinal orientations about racial bias, discrimination, and inequality. However, exposure to credible information about implicit bias (vs. automaticity) strengthened the effect of social dominance orientation, system justification, and explicit black attitudes on endorsement of racial stereotypes. People who believe in the legitimacy of social hierarchy, perceive the status quo as fair and desirable, or hold explicitly negative attitudes towards blacks were more motivated to attribute the causes of racial inequality in the criminal justice system, employment decisions, and socio-economic status to personal (vs. situational) factors. This finding highlights both the ability for these attitudinal orientations to resist falsification and assimilate new information in a confirmatory way, as well as a potential pitfall associated with efforts to educate the public about implicit racial bias and its consequences. While individuals characterized by these beliefs were more inclined to justify and rationalize inequality upon learning about the existence of racial bias, it is important to note that, independent of condition, these individual differences still predicted the motivated rejection of implicit racial bias, decreased prejudice regulation, and opposition to public policy intended to help minorities.

Together, the results from three experiments indicates that exposing White Americans to information about implicit racial bias *can* increase self-perceived implicit bias, but, most critically, may also promote unfavorable opinions of social scientists and increase the effect of socio-political attitudes on endorsement of racial stereotypes. Belief in prejudice and discrimination appears to be unaffected by this information, suggesting that these beliefs may be relatively more stable and resistant to new information than self-perceived implicit bias or attitudes towards social science. Still, the belief in the existence

of racial prejudice and self-perceived implicit racial bias predicted an increase in prejudice regulation, regardless of condition, although the former was a more reliable predictor of these outcomes. Similarly, belief in racial prejudice and, to a lesser extent, perceptions of social science each mediated the relationship between 1) explicit racial attitudes, sociopolitical orientations, and egalitarian motivations and 2) prejudice-regulation and endorsement of racial stereotypes. Self-perceived implicit bias did not mediate these relationships. People who hold explicitly negative attitudes towards blacks or who are motivated to justify the legitimacy of social hierarchy or the status quo were less willing to believe in racial prejudice and held less favorable attitudes towards social science, which increased racial stereotyping.

This overall pattern demonstrates that the motivated acceptance or rejection of implicit racial bias is an important determinant of prejudice-regulation, stereotyping, and public policy attitudes. Further, these motivated reasoning processes are clearly linked to theoretically relevant individual differences in sociopolitical orientations, explicit racial attitudes, and egalitarian motivations. However, efforts to communicate the science of implicit racial bias, while capable of increasing self-awareness of one's own racial bias, nonetheless risks backlash against social scientists, and can increase the endorsement of racial stereotypes among individuals who are motivated to reject such evidence and justify inequality. Importantly, these motivated reasoning processes have serious real-world implications. For example, endorsement of racial stereotypes undermined support for public policy intended to remediate racial inequality, minimize discrimination, and improve the lives of racial minorities.

The failure to observe any reliable effect of the experimental manipulation on beliefs in racial prejudice and discrimination highlights a potential disconnect between these beliefs and self-perceived implicit bias, which was indeed influenced by the presentation of credible information on the scientific literature in this domain. This possibility is perhaps further punctuated by the diverging effects of sociopolitical orientations, explicit racial attitudes, and egalitarian motivations for self-perceived implicit bias and the belief in prejudice and discrimination. That is, people characterized by negative explicit racial attitudes and a belief in the legitimacy of social hierarchy and the status quo were able to perceive their own implicit racial bias, whereas individuals intrinsically motivated by egalitarian ideals were unwilling or unable to do so. Individuals characterized by the former may be able and willing to recognize their own unfavorable attitudes towards blacks, but because they regard these beliefs as an accurate representation of the social world, they appear more willing to endorse stereotypes and less willing to regulate their prejudices. In contrast, intrinsic egalitarians may be more motivated to engage in prejudice-regulation, but because these values are so closely embedded in their self-concepts, these individuals may also struggle to recognize that, despite their best efforts, their unconscious attitudes are nonetheless racially biased.

More generally, the upshot of this pattern of evidence is that White people struggle to reliably connect their own individual-level biases to aggregate-level inequalities. For example, it appears that the ability to recognize and admit one's own implicit racial bias is more common among extrinsic egalitarians and individuals who are motivated to justify and rationalize existing inequality as legitimate and fair. None of these characteristics were consequential for the motivation to control prejudiced

reactions. Alternatively, people who are intrinsically committed to egalitarian ideals may be able and willing to recognize prejudice and discrimination in society, but may still struggle to perceive or accept their own implicit racial bias or may be otherwise unwilling to make that inference on the basis of the information presented to them in this context. Additional research that provides individuals with *direct* feedback about personal implicit bias, and directly communicates the link between such bias and racial discrimination and inequality in society, is needed to help clarify this dynamic. Study 2 was, in part, undertaken for this reason.

Pilot 2: Allegation of Implicit Racial Bias

The experimental paradigm used in Study 2 was first pilot tested on a sample drawn from the target population. Pilot 2 employed a single independent variable design (Implicit Racial Bias Feedback vs. Egalitarian Feedback vs. No Feedback). The primary goal of the pilot for Study 2 was to independently validate the experimental stimuli. For example, Pilot 2 will provide preliminary evidence of the affective and cognitive consequences of the racial bias feedback, that such feedback is perceived as believable, and that participants do not suspect an intention to deceive or mislead them. A major concern is that evidence indicating that people are motivated to reject evidence of implicit bias feedback could be a form of reactance to a perceived intention of the researchers to deceive or mislead participants with fictitious information. The pilot study is intended to empirically assess these concerns. In general, no differences between conditions are expected for the perceived believability or comprehensibility of the message prompt. However, it is expected that participants who receive feedback suggesting that they harbor implicit bias will report higher levels of negative affect and

lower levels of positive affect than participants in the no feedback condition. Lastly, another major goal of this pilot study is to evaluate the effect of the IAT feedback and motivated reasoning processes on the availability of cognitive resources (i.e., Stroop task). I expect that the response latency of individuals who score higher on the full motivated reasoning battery (described below) to be slower for incongruent vs. congruent trials during the Stroop test.

Method

Participants

Participants were 368 White U.S. citizens recruited from Amazon MTurk (63% females, 37% males; mean age = 37.44, *SD* = 12.77). Most participants were modestly affluent (44% report a family income greater than 50K) and educated (47.3% have earned at least a Bachelor's degree). With this sample size, to detect mean level differences between the experimental and control group, I estimated that I had 47% power to detect a Cohen's *d* of 0.2, 99% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit White U.S. citizens and would compensate participants $0.50 for their time. The name of the study is intended to increase the expectation that one's beliefs and attitudes about other people would be directly measured.

Participants first viewed a consent form for the study, and were then randomly assigned to the bias feedback, egalitarian, or no feedback (i.e., control) condition. Before

completing the Race IAT, participants in the control condition received the following instructions:

In this study, you will be shown pictures of several individuals, and will be asked to pair each of these pictures with a list of words. Please remember that we are interested in your perspective. There are no right or wrong answers to the questions, and your first response is usually the best. After completing the task you will be asked to provide some information on your opinion about current events and your basic demographics information. This is a challenging task, but it's necessary for the aim of this study. Please try hard to help us in our analysis.

Please click next to proceed to the test.

After completing the Race IAT, participants in the no feedback (i.e., control) condition receive the following feedback:

Thank you for completing our sorting task. Next, you will be asked to provide some information on your opinion about current events and your basic demographics information. Please click next to proceed to additional questions.

Before completing the Race IAT, participants in the bias and egalitarian feedback conditions received the following instructions:

The following test compares your attitudes toward two different racial groups. It is a measure of racial attitudes. In particular, this test was designed to reveal your unconscious beliefs and attitudes towards racial minorities. In this experiment, we are interested in measuring your unconscious racial attitudes toward Blacks and Whites as accurately as possible. This test will reveal your unconscious beliefs with a high degree of accuracy and precision- even if you are not aware of these beliefs and even if these beliefs are inconsistent with your conscious beliefs and values. This is a challenging task, but it's necessary for the aim of this study. Please try hard to help us in our analysis of individuals' racial attitudes.

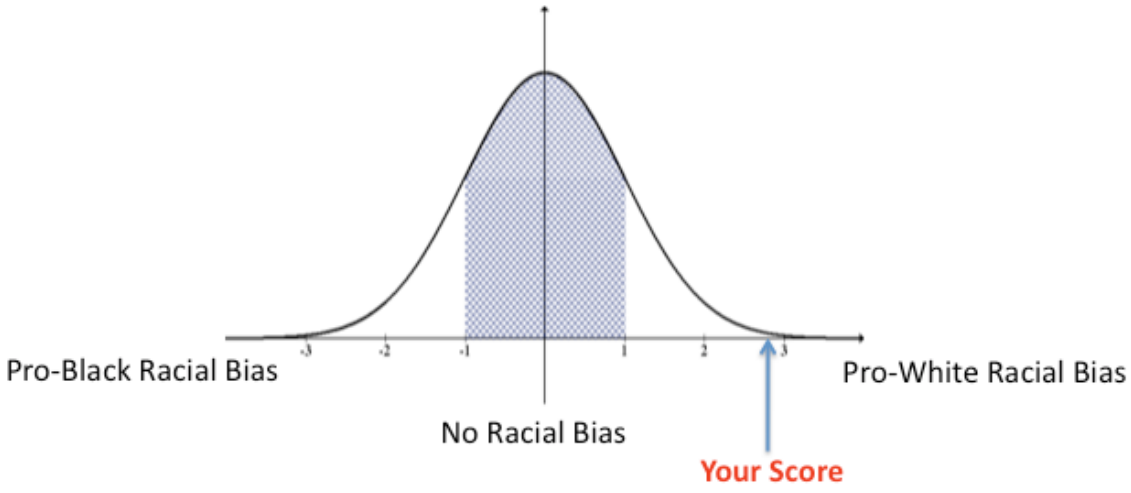
Please click next to proceed to the test of unconscious racial attitudes.

These instructions were intended to increase the belief in the diagnosticity of the Race IAT. After completing the Race IAT, participants in the bias feedback condition received the following information and graph:

Based on this test, it is clear that you are racially biased at the unconscious level. More specifically, the results indicate that you have a strong preference in favor of White people compared to Black people. Below, you'll see a graphical representation of where your tests scores place you compared to the population.

Please note that a significant amount of research in the social and behavioral indicates that scores on this test predict people's judgment of and behavior towards racial minorities. For example, one study demonstrated that people with racial bias often avoid being friends with racial minorities or treat them unfairly in social situations, even though they didn't intend to. This means that there is a really good chance that you harbor unconscious bias towards Black people and are particularly likely to discriminate against them.

Below is a graphical representation of your estimated performance on this test of unconscious racial bias. The figure is not intended to be a perfect representation of your score, but to give you a general sense of how your unconscious racial attitudes compare to other people in the population.

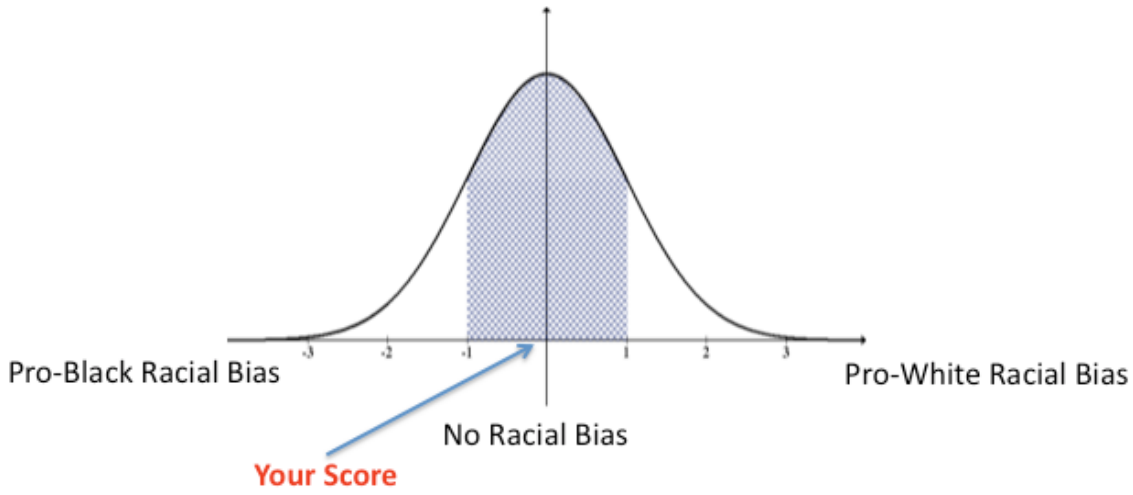


After completing the Race IAT, participants in the egalitarian feedback condition received the following information and graph:

Based on this test, it is clear that you are racially fair at the unconscious level. More specifically, the results indicate that you have an equal preference for White people and Black people.

Please note that a significant amount of research in the psychological sciences indicates that scores on this test predict people's judgment of and behavior towards racial minorities. For example, one study demonstrated that White employers who scored lower on measures of unconscious bias evaluated White job applicants and Black job applicants fairly. This means that there is a really good chance that you do NOT harbor unconscious bias towards Black people and are particularly likely to treat people of different races fairly.

Below is a graphical representation of your estimated performance on this test of unconscious racial bias. The figure is not intended to be a perfect representation of your score, but to give you a general sense of how your unconscious racial attitudes compare to other people in the population.



Next, all participants proceeded to complete measures of affect, manipulation checks, motivated reasoning, and the Stroop task. All of these measures are described below. The Stroop task followed the same procedure as that used by Richeson and Trawalter (2005). Specifically, participants will report, as quickly as possible, the correct color of the font for the stimulus words. By pressing the appropriate response option on the computer screen, participants either responded to words that are 1) itself the name of

a color (incompatible trials) different from the color of its font, or 2) nonsensical letter strings (control trials). The presentation order of the color and control words was randomly determined, but an equal number of each per trial was presented. Color names or the control appeared sequentially, in red, yellow, green or blue (randomly determined). Each stimulus word appeared for a maximum 2,000 ms, separated by intervals of 1500 ms. The task consisted of 24 practice trials followed by 6 blocks of 12 trials each, for a total of 72 experimental trials. Response latency to each stimuli was measured, and used to compute measures of cognitive resources (slower response time to incompatible vs. control trials was taken as an indicator of cognitive depletion).

Finally, participants completed measures of demographics and were fully debriefed.

Measures

Manipulation Checks. 5 items were used to validate the manipulation. On a 7-point scale ranging from 1 (”Strongly Disagree”) to 7 (”Strongly Agree”), participants indicated whether 1) “The information presented to me was believable” and 2) “I wondered if the information presented to me was accurate”. Participants also reported their confidence in each judgment. Perceived believability was indexed by the average of responses to the item and expressed confidence in that response (M=.70, SD=.62). Perceived accuracy was indexed the same way (M=.60, SD=.45). Participants also indicated if they thought “the researcher intentionally tried to deceive or mislead you” (“No”=0, “Yes”=1; M=.40, SD=.49).

Race IAT. The most commonly used method to measure implicit attitudes is the computer-administered categorization task, the Implicit Association Test (Greenwald,

McGhee, and Schwartz 1998). IAT scores were computed using the most recent algorithm from Greenwald, Nosek, and Banaji (2003). Higher scores represent stronger associations between White-positive, and lower scores represent stronger associations between Black-positive (M=.60, SD=.10).

Affect. 31 items were used to measure affect, following the procedure described by prior research on prejudice-regulation (e.g., (Devine et al., 1991; Monteith & Voils, 1998; Monteith, Voils, Ashburn-Nardo, 2001), to form measures of positive affect (M=.57, SD=.24, Cronbach's alpha=.92), self-directed negative affect (M=.20, SD=.24, Cronbach's alpha=.95), other-directed negative affect (M=.18, SD=.22, Cronbach's alpha=.88), depressed (M=.22, SD=.29, Cronbach's alpha=.96), and discomfort (M=.20, SD=.23, Cronbach's alpha=.94). Taking the mean of all measures of negative affect, I also computed an overall negative affect measure (M=.20, SD=.22, Cronbach's alpha=.97).

Full Motivated Reasoning Battery (FMR-B). 27 items were used to measure the motivated rejection of personal and general implicit racial bias, prejudice, and discrimination, concerns about implicit bias, willingness to participate in an anti-bias intervention, and perceptions of the credibility, validity, and objectivity of the IAT and of social scientists, more generally, and acceptance of the results of the IAT[3]. Responses across all 27 items were average to obtain a general index of motivate reasoning in this context (M=.55, SD=.15, Cronbach's alpha=.87), such that lower values represent greater motivated reasoning. However, 4 theoretically distinct subscales were also obtained, and

[3] The item used to measure acceptance of the IAT results was only administered to participants who were assigned to a feedback condition. For participants in the control condition, this item was not included in computations of FMR-B.

allow for a more nuanced investigation of the specific beliefs relevant to these motivated reasoning processes. Accordingly, these subscales are the focus of my analyses. Each subscale is described below.

Self-Perceived Implicit Racial Bias. Participants reported a belief in their own implicit racial bias across 6-items. On a 7-point scale, participants responded to such items as, "How likely is it that your unconscious beliefs are unfavorable toward racial minorities?", "Do you believe that your unconscious racial attitudes influence your behavior towards racial minorities in an unfair way?", and "How worried are you that you are unconsciously prejudiced towards racial minorities?". Higher values represent increased belief in one's implicit racial bias (M=.31, SD=.23, Cronbach's alpha=.90).

Perceptions of Social Science. Participants reported their belief in the credibility, objectivity, integrity, and extremism of social science and scientists across 8-items. On a 7-point scale, participants responded to such items as, "To what extent are social scientists who study the psychology of unconscious racial bias motivated by a political or ideological agenda?", "How credible are social scientists who study the psychology of unconscious racial bias?", and "Social scientists think that all White people are racists.". Higher values represent more favorable attitudes towards social science and scientists (M=.39, SD=.21, Cronbach's alpha=.87).

Belief in Racial Prejudice and Discrimination (some items adapted from O'Brien et al., 2010). Participants reported their belief in the existence of implicit racial bias, and racial prejudice, discrimination, and inequality in the general population across 7 items on a 7-point scale. Such items include 1) " How common is unconscious racial prejudice in America?", 2) "How likely is it that unconscious racial attitudes biases people's

judgments and behavior towards racial minorities?", 3) "Do you think differences between racial groups can be explained by the effects of unconscious racial bias and prejudice?", 4) "How common is racial prejudice in America?", and 5) "Do you think that racial minorities are held back in society because of racial prejudice?". Higher values represent greater belief in the existence of implicit racial bias and prejudice in the general population (M= .61, SD=.21, Cronbach's alpha=.91).

Perception of the IAT and Agreement With Test Results. Participants reported their belief in the validity, credibility, and objectivity of the IAT 4 items on a 7-point scale. Such items include 1) "In your opinion, how credible is this test?", 2) "In your opinion, how objective is this test?", 3) "In your opinion, how valid are the results of this test? 4) "In your opinion, how useful is this test for understanding people's racial attitudes?" Higher values represent increased perceptions of the credibility and validity of IAT (M= .48, SD=.28, Cronbach's alpha=.93). Furthermore, participants who received implicit bias feedback, were reported their agreement with the test results using the following item "To what extent do you agree with your results from the test?" on a 7-point scale, with higher values representing higher levels of agreement with the test results (M= .50, SD=.37).

Stroop Task. A difference score between response latencies for incongruent and congruent trials was computed as an indicator of cognitive depletion. Higher values represent slower responding to incongruent (vs. congruent) trials and thus higher levels of cognitive depletion (*M*=5.26, *SD=.94*). Unlike all variables used in these analyses, transformed scores were not recode to run from 0-1.

Demographics. Participants reported their age, gender, race, family income, and

level of education.

See Table 2.1 and 2.2 for means, SD, alphas, and correlations between all measures used in analyses for Pilot 2.

Results

Two dummy-coded variables were created to represent condition assignment, with the bias feedback condition as the reference group. For analyses comparing differences between the egalitarian feedback and control groups, an additional set of dummy-coded variables were created such that the control condition serves as the reference group. For all of the models used in analyses reported below, both dummy-coded variables are included along with covariates.

Manipulation Check

Participants' judgments of the believability and accuracy of the information contained in the experiment did not differ between the bias feedback and the no feedback control group (p>.3). However, participants in the bias (vs. egalitarian feedback) condition judged the information as less believable *(b* = .20 (*95% CI = .01, .39) p* =.039) but similar in perceived accuracy *(p>.75).* Participants were also more likely to report the use of deception in the bias feedback condition compared to the egalitarian feedback *(Exp(b)* = .27 (*95% CI = .15, .46*), *p* <.001) and no feedback condition *(Exp(b)* = .10 (*95% CI = .05, .19*), *p* <.001). Furthermore, participants in the egalitarian feedback (vs. control) condition regarded the information as more believable *(b* = .14 (*95% CI* = -*.01, .28) p* =.06), more accurate *(b* = -.07 (*95% CI* = -*.12, -.02) p* <.01), and less deceptive *(Exp(b)* = 2.65 (*95% CI = 1.45, 4.87*), *p* <.01).

Thus, participants regarded the bias feedback condition as comparable in terms of perceived accuracy and believability, but more deceptive, than the control condition. However, the egalitarian feedback was perceived as more believable, more accurate, and less deceptive than the control group, and more believable than the bias feedback condition. It is noteworthy that the perceived believability and accuracy of the information did not differ between participants in the bias and no feedback condition, whereas the same exact feedback stimuli informing participants that they do not harbor implicit bias was judged as more believable and accurate relative to the control group. This pattern of results suggests that participants were particularly motivated to regard the egalitarian feedback as an accurate and credible characterization of their implicit racial attitudes.

Importantly, that I observed no differences in the perceived believability and accuracy of the information contained in the bias feedback compared to the control group supports the appropriateness of using these stimuli for Study 2. However, and perhaps not surprisingly, participants judged the information as more deceptive in the bias (vs. no feedback) condition, suggesting that participants held some degree of suspicion that implicit bias feedback was intended to deceive them. For this reason, I turn to the open-ended responses to better ascertain whether participants in the bias feedback condition regarded the information as fabricated and intentionally designed to deceive them into believing, incorrectly, that they harbor implicit racial bias.

Participants were asked to describe what they believed was the true purpose of the experiment. Each open-ended response were content coded by two independent raters for several themes, including 1) general belief in the use of deception and fabrication, 2)

whether the purpose of the study was to present information about race and prejudice to inform people and learn about racial biases and attitudes, or 3) whether the purpose of the study was to persuade, deceive, or influence opinions about racism, personal bias, or white privilege.

Complete information about the coding scheme and the instructions provided to the raters is available in the Appendix. The presence of each category in the open-ended response was coded as 1, and its absence was coded as a 0. Inter-rater agreement across categories ranged from low (.37) to acceptable (.57), but was altogether unsatisfactory (see Table 2.3 for kappa across each category). A higher level of agreement is desirable to better understand how open-ended responses varied across condition. Thus, both coders convened to attain consensus on points of disagreement. I rely upon these consensus judgments for the analyses reported below.

Ratings of the open-ended responses were regressed on the dummy-coded variables for condition assignment. Consistent with the results of the manipulation check items, participants in the bias feedback condition were more likely to report a general belief in the use of deception, compared to participants in the no feedback condition *(Exp(b)* = 5.24 (*95% CI = 1.47, 18.73), p* <.05). Participants in the bias feedback condition were more likely to report that the purpose of the study was to inform people and learn about racial bias than participants in the no feedback condition *(Exp(b)* = 1.65 (*95% CI = .98, .2.78), p* =.06). Participants in the egalitarian feedback condition were also more likely to report that the purpose of the study was to inform people and learn about racial bias than participants in the no feedback condition *(Exp(b)* = 1.99 (*95% CI = 1.17, 3.36) p* =.01), but did not differ from the bias feedback condition in this regard.

More critically, no significant differences across condition emerged in the extent to which participants believed that the purpose of the study was to deceive or manipulate people's opinions with fabricated information and feedback about race. No other comparisons obtained significance.

This pattern of results from open-ended responses suggests that participants who received *any* kind of feedback were more likely to report that the study was intended to inform people about race and to learn about racial bias. Furthermore, and consistent with the results from the close-ended manipulation check items, participants were particularly more inclined to believe that deception was used in the bias feedback, compared to the egalitarian feedback condition. It is notable that perception of deception did not differ between the no feedback and egalitarian feedback, despite the fact that the only difference between the egalitarian and bias feedback was the direction of that feedback (i.e., same presentation, information, and diagram). However, and perhaps of central importance to the validity of the experimental paradigm, this increased perception of deception did not lead participants to infer that the information contained in the bias feedback condition was fabricated and intentionally designed to persuade, deceive, or influence their opinions about race, bias, or white privilege.

Together, the results from close-ended and open-ended manipulation check items reveal that participants were particularly motivated to believe that the egalitarian feedback was accurate and objective, and were motivated to believe that the bias feedback was deceptive. Because bias feedback is expected to be inherently threatening (and egalitarian feedback inherently affirming), it is no surprise that people were motivated to reject or accept it. Indeed, these findings are consistent with the proposition

that people engage in motivated reasoning to maintain egalitarian self-images in response to feedback about implicit bias. However, I find no evidence to suggest that participants' belief in the use of deception in the bias feedback condition was based on the belief that such information was fabricated or intentionally designed to mislead or influence their beliefs about race, bias, or privilege. For these reasons, this experimental paradigm is appropriate for use in Study 2.

Main Effects of Experimental Manipulation on Affect

Next, I examined the effect of the experimental manipulation on affect. Marginal means for each measure of affect, separate for feedback condition, are represented in Figure 2.1. Results indicate that participants in the bias feedback condition, compared to the no feedback (and egalitarian feedback; ps<.05) conditions, reported significantly lower levels of positive affect *(b* = .06 (*95% CI = .003, .12) p* =.-39), and significantly higher levels of negative affect overall *(b* = -.07 (*95% CI = -.12, -.02) p* =.012), depressed affect *(b* = -.07 (*95% CI = -.14, .001) p* =.052), discomfort *(b* = -.07 (*95% CI = -.12, -.01) p* =.022), and self-directed negative affect *(b* = -.09 (*95% CI = -.16, -.03) p* =.003), but did not report any difference in other-directed affect (p>.15). No significant differences in affect were observed between participants in the egalitarian and no feedback condition, although participants in the former reported significantly higher levels of positive affect *(b* = -.06 (*95% CI = .00, .11) p* =.051). Thus, bias feedback induced negative affect and reduced positive affect, relative to both egalitarian and no feedback conditions.

Main Effects of Experimental Manipulation on Motivated Reasoning

The effect of experimental condition on each indicator of motivated reasoning is represented in Figure 2.2. These results indicate that participants in the bias feedback condition, compared to both the egalitarian *(b* = .22 (*95% CI = .17, .26) p* <.001) and no feedback condition *(b* = .15 (*95% CI = .10, .20) p* <.001), were less likely to regard the IAT as a credible, valid, and objective measure of their implicit attitudes. However, no other significant differences in motivating reasoning were observed between participants in the bias and no feedback conditions.

In contrast, participants in the egalitarian (vs. bias) feedback condition were also more likely to agree with the results of the IAT *(b* = 4.25 (*95% CI = .58, 7.91) p* =.023) and to regard the IAT (*b* = .22 (*95% CI = .17, .26) p* <.001) and social scientists as credible, objective, and apolitical *(b* = .05 (*95% CI = .00, .04) p* <.059), but less likely to perceive their own bias *(b* = -.07 (*95% CI = -.13, -.01) p* =.023). Compared to the control group, participants in the egalitarian feedback condition had more favorable attitudes towards social scientists (*b* = .06 (*95% CI = .01, .11) p* =.027) and the IAT *(b* = .06 (*95% CI = .01, .11) p* =.011). Interestingly, participants who received egalitarian feedback (vs. no feedback) were, as expected, less likely to perceive their own implicit bias *(b* = -.06 (*95% CI = -.12, -.01) p* =.029), but *more* willing to recognize the existence of prejudice and discrimination in society *(b* = .06 (*95% CI = .01, .11) p* =.02). Perhaps feedback that was affirming of one's egalitarian self-concept made participants more comfortable and willing to perceive and recognize discrimination in society, even if such feedback also reduces self-awareness of one's own bias.

Together, these results indicate that participants who received bias feedback were particularly motivated to discredit and devalue the IAT. However, participants who

received egalitarian feedback were particularly motivated to view the IAT and social scientists, more generally, as credible and objective. Furthermore, egalitarian feedback, but not bias feedback, directly impacted their beliefs about implicit bias and racial discrimination, such that participants who received this feedback were less willing to perceive their own bias but more willing to believe in the existence of discrimination and prejudice in the general population.

Relationship Between Affect and Motivated Reasoning

Figure 2.3 represents a coefficient plot for the estimated relationship of each affect dimension and each indicator of motivated reasoning. These results indicate that overall negative affect *(b* = .32 (*95% CI = .19, .45) p* <.001), self-directed negative affect (*b* = .24 (*95% CI =.13, .35) p* <.001), other-directed negative affect *(b* = .24 (*95% CI =.10, .38) p* =.001), depression *(b* = .20 (*95% CI = .10, .31) p* <.001), and discomfort *(b* = .32 (*95% CI = .19, .45) p* <.001) were associated with self-perceived implicit bias- participants who experience negative affect were *more* likely to report their own bias. Positive affect, but no other affective measure, was associated with more favorable perceptions of the IAT (*b* = .14 (*95% CI = .03, .25) p* <.001). Finally, while only other-directed negative affect associated with less favorable attitudes towards social scientists *(b* = -.10 (*95% CI = -.21, .01) p* =.089), all other negative affect measures were associated with an increased belief in the existence of prejudice and discrimination in society (ps<.001). Thus, negative affect, which can be induced by bias feedback, appears able to increase both self-perceived bias and the belief in the existence of prejudice and discrimination in society.

Affect as a Mediator for Effect of Implicit Bias on Motivated Reasoning

Because 1) implicit bias feedback induced negative affect and undermined positive affect, and because 2) the former was associated with self-perceived implicit bias and the latter was associated with perceptions of the IAT and acceptance of the IAT results, I conducted a series of mediation analyses to better understand the relationship between implicit racial bias feedback, affect, and motivated reasoning, with the expectation that negative affect would mediate the relationship between feedback and self-perceived bias. For this analysis, I examine the mediating role of overall positive or overall negative affect for the effect of receiving implicit racial bias feedback (coded as 0 vs. control coded as 1) on each indicator of the FMR-B. To do so, I ran a bootstrap analysis that tested for the indirect effect of each predictor on the dependent variable, retaining all covariates.

The results of this analysis indicate that negative affect was a critical mediator for the effects of implicit bias feedback on self-perceived bias. Specifically, with overall negative affect submitted as a mediator, I observe a significant indirect effect of implicit racial bias feedback on self-perceived implicit bias *(b = -.02, SE = .01, (95% CI = -.03, .001), p*=.039*)* but not belief in the existence of prejudice and discrimination (p>.15), perceptions of the IAT *(p*>.15*)* or perceptions of social scientists *(p*>.25*)*. Thus, the experience of negative affect following implicit bias feedback appears to be an important mediator for the effect of that feedback on self-perceived bias.

Effect of Motivated Reasoning on Cognitive Depletion

The difference in response latency for incongruent and congruent trials were regressed on the FMR-B. These results indicate that motivated reasoning was associated

with cognitive depletion (i.e., slower response latency to incongruent vs. congruent Stroop trials; *(b* = -161.31 (*95% CI =-312.54, -10.08*), *p* =.037)

Discussion

Implicit bias feedback undermined positive affect, induced negative affect, and decreased perceptions of the IAT as a credible, valid, and objective measure of one's attitudes, relative to both egalitarian and no feedback conditions. In contrast, egalitarian feedback decreased self-perceived bias, but increased positive affect, perceptions of the credibility and validity of the IAT and of social scientists, more generally, relative to both the bias and no feedback condition. Affect was also closely related to motivated reasoning, such that people who felt negatively (positively) following completion of the IAT were more (less) willing to engage in the motivated rejection of implicit racial bias. Indeed, negative affect mediated the relationship between implicit bias feedback and self-perceived implicit bias. Finally, this motivated rejection of implicit racial bias was cognitively depleting, as indexed by the Stroop test, suggesting that maintaining one's egalitarian self-image in the face of implicit bias feedback requires the expenditure of cognitive effort.

More importantly for the purposes of Pilot 2, the experimental stimuli are appropriate for use in Study 2. While participants were more inclined to believe that deception was used in the bias feedback, and that egalitarian feedback was an accurate and credible characterization of their implicit racial attitudes, I observed no differences in the perceived believability and accuracy of the information contained in the bias feedback compared to the control group. Furthermore, the increased perception of deception in the biased feedback condition did not lead participants to infer that the information was

fabricated and intentionally designed to mislead people about their attitudes. For these reasons, I have sufficient confidence in the validity of the experimental paradigm used in Study 2.

Study 2: Allegation of Implicit Racial Bias

In Study 2, I experimentally manipulate participants' belief in their own implicit bias using a false-feedback paradigm following completion of the Implicit Association Test (Greenwald, McGhee, & Schwartz 1998). This study employed a single independent variable design (IAT Feedback: Racial Bias, Accurate, Control). The purpose of Study 2 was to evaluate the psychological factors associated with the motivated rejection of alleged implicit racial bias, and the consequences of these motivated reasoning processes for outcomes of interest. In particular, I seek to replicate the correlational results of Study 1 (H1; H6-H11), suggesting that sociopolitical orientations, egalitarian motivations, and racial attitudes are associated with the motivated rejection of implicit racial bias, and that these mediated the effect of individual differences on prejudice-regulation and racial stereotyping. I also test the hypotheses that feedback that one harbors implicit racial bias will strengthen or activate the effect of sociopolitical orientations, racial attitudes, and egalitarian motivations on the motivated rejection of all evidence of racial discrimination and inequality and implicit racial bias (H2-3). Finally, I also tested the hypotheses intrinsic (vs. extrinsic) egalitarians will be more accepting of evidence of their own implicit bias (H4), especially if there is a high level of dissociation between their implicit and explicit racial attitudes (H5).

Study 2, unlike pilot 2, also included an experimental condition in which participants received accurate feedback, based on their performance on the IAT. The

inclusion of accurate feedback condition allowed me to control for a potential confound in the racial bias feedback. Specifically, it is possible, albeit unlikely, that some individuals' have accurate introspective awareness of their implicit racial attitudes. As a result, these individuals may be motivated to reject evidence of their own implicit racial bias, but not because of the threat to their unprejudiced self-images, as expected. Instead, to the extent that these participants have accurate introspective awareness of their implicit attitudes, incongruent feedback may undermine their belief in the credibility of the test. Thus, some may be motivated to reject implicit bias feedback because it is inconsistent with their "accurate" self-knowledge, not because it threatens their self-concepts. However, because the majority of the sample is expected to have pro-white implicit bias, only some participants in the implicit bias condition will receive feedback incongruent with "accurate" self-knowledge that they do *not* harbor implicit bias. Comparisons between the implicit bias condition and participants who receive implicit feedback in the accurate feedback condition will allow for an assessment of the role of "accurate" self-knowledge for understanding the consequences of receiving feedback suggesting that one is implicitly biased.

Method

Participants

Participants were 444 White U.S. citizens recruited from Amazon MTurk (52% females, 47% males; mean age = 36.15, *SD* = 12.19). Most participants were modestly affluent (43% report a family income greater than 50K) and educated (48% have earned at least a Bachelor's degree). Of these individuals, 277 were retained at T2 (62%). With this sample size, to detect mean level differences between the experimental and control

group, I estimated that I had 39% power to detect a Cohen's *d* of 0.2, 96% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit white U.S. citizens for a two-part study, and would compensate participants $1.50 total for their time (T1=$0.25, T2=$1.25).

This study used a 2-wave panel design. Pre-manipulation measures were assessed at Time 1 (T1). Participants first viewed a consent form, and complete the pre-manipulation battery. Up to one week later, participants were contacted to complete the second part of the study. At Time 2 (T2), participants were randomly assigned to the racial bias feedback, accurate bias feedback, or control condition (described in pilot 2). The accurate feedback condition provided participants with correct information about their performance on the Race IAT. More specifically, in the accurate feedback condition, participant received the same pre-IAT instructions as participants in the race bias condition. Participants who scored above 0 on the Race IAT, indicating pro-white bias, received the same post-IAT feedback as participants in the bias feedback condition. However, participants who scored below 0, indicating a pro-black bias, received the egalitarian feedback.

Measures

See Table 2.4 for means, SD, and correlations between all measures used in analyses for Study 2. The T1 and T2 measures were identical to that used in Study 1. However, Study 2 also utilized the affect measures, Race IAT, and FMR-B described in Pilot 2. Furthermore, taking the absolute value of the difference in scores between the

Race IAT and attitudes towards blacks, a measure of implicit-explicit dissociations was obtained, such that higher values represent increased levels of implicit-explicit dissociation (M=.17, SD=.13).

Results

Two dummy-coded variables were created to represent condition assignment, with the control condition as the reference group. For analyses comparing differences between the accuracy feedback and bias feedback, an additional set of dummy-coded variables was created such that the bias feedback condition serves as the reference group. Only participants in the accuracy feedback condition that received bias feedback are included in the analyses described below. Interaction terms were constructed by taking the product between the dummy-coded condition variable and rescaled continuous predictor variables. For moderation analyses, I also examine the effect of the moderating variable separately for each condition when simple slopes analyses fail to clarify the nature of the interaction. All statistical models include age, gender, income, and trust in social science as covariates.

Main Effects of Experimental Manipulation on Affect

Next, I examined the effect of the experimental manipulation on affect. Marginal means for each measure of affect, separate for feedback condition, are represented in Figure 2.4. Results indicate that participants in the bias feedback condition, compared to the no feedback conditions, reported significantly lower levels of positive affect *(b* = -.15 (*95% CI = -.23, -.08) p* <.001), and significantly higher levels of negative affect overall *(b* = .07 (*95% CI = .01, .14) p* =.024), discomfort *(b* = .09 (*95% CI = .02, .15) p* =.002), depression *(b* = .07 (*95% CI = -.01, .15) p* =.097), and self-directed negative affect *(b* =

.11 (*95% CI = .03, .18) p* =.007), but did not report any difference in other-directed affect. No significant differences in affect were observed between participants in the accuracy and bias feedback condition, with one major exception. Participants in the accuracy feedback condition experienced significantly more other-directed negative affect than participants in the bias and no feedback conditions *(b* = .09 (*95% CI = .003, .17) p* =.043).

Effects of T1 Individual Differences x Message Condition on Affect

It was hypothesized that feelings of guilt and shame among low prejudiced, intrinsic egalitarians would attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation. The first step to test this hypothesis is to examine whether experimental condition interacted with intrinsic and extrinsic egalitarian motivations, separately, to predict affect. Evidence in support of this hypothesis would be indicated by a significant interaction between intrinsic (but not extrinsic) egalitarianism and the bias feedback condition to predict negative affect. A similar set of analyses was conducted with SDO, system justification, and implicit and explicit racial attitudes to explore their role as moderators for the effect of feedback on affect. Below I report the interaction and simple slopes analyses for significant models only.

The interaction between SDO and the accurate (vs. control) condition obtained significance for negative overall (b=-.42, CI(-.81, -.03), p<.05), self-directed negative affect (b= -.48, CI(-.93, -.03), p<.05), and discomfort (b=-.50, CI(-.91, -.08), p<.05). At 1 SD above the mean of SDO, the accurate (vs control) feedback increased overall negative affect (b=.15, CI(.07, .23), p<.001), self-directed negative affect (b=.18, CI(.08, .27),

p<.001), and discomfort (b=.17, CI(.08, .25), p<.001). Similarly, at 1 SD below the mean of SDO, the accurate (vs. control) condition increased overall negative affect (b=.31, CI(.11, .51), p=.002), self-directed negative affect (b=.36, CI(.13, .59), p=.002), and discomfort (b=.36, CI(.15, .57), p=.001). While the difference between the accuracy and control condition obtained significance for people both high and low in SDO, inspection of the unstandardized coefficients indicates that accurate bias feedback more strongly induced negative affect among people low (vs. high) SDO. Consistent with this, only in the accurate feedback condition was SDO marginally significant associated with overall negative affect (b=-.29, CI(-.64, .06), p=.099), self-directed affect (b=-.36, CI(-.78, .05), p=.08), and discomfort (b=-.34, CI(-.73, .05), p=.087). Similarly, the interaction between SDO and the bias (vs. control) condition obtained significance for positive affect (b=.42, CI(.04, .81), p<.05). At 1 SD above the mean of SDO, (b=.15, CI(.07, .23), p<.001) and below the mean of SDO (b=-.35, CI(-.54, -.15), p<.001) the bias (vs control) feedback decreased positive affect. Furthermore, the relationship between SDO and positive affect only obtained significance in the bias feedback condition. (b=.51, CI(.18, .83), p=.002), suggesting that bias feedback was particularly likely to undermine positive affect among individuals low in SDO. The relationship between SDO and each of these measures of affect, for each experimental condition, are represented in Figures 2.5.

The 2-way interaction between feedback condition and egalitarian motivations did not obtain significance for affect. To better understand the role of egalitarian motivations in conditioning the impact of implicit bias feedback, I tested the 3-way interaction on affect between 1) egalitarian motivations, 2) experimental condition, and 3) implicit-

explicit dissociations. These analyses failed to reveal a significant 3-way interaction on any measure of affect.

Affect and Motivated Reasoning

To evaluate the extent to which the experience of positive or negative affect was associated with motivated reasoning, each measure of affect was regressed separately on each indicator of racial attitudes, sociopolitical orientations, and egalitarian motivations. Figure 2.6 provide a coefficients plot that graphically represent the unstandardized coefficients and 95% CI for each measure of affect from models estimating its impact on self-perceived implicit racial bias, perceptions of social scientists, perceptions of the IAT, agreement with test results (feedback conditions only) and belief in prejudice and discrimination, respectively. These results indicate that overall negative affect (b=.14, CI(.05, .23), p=.003), self-directed negative affect (b=.13, CI(.05, .20), p=.001), depressed affect (b=.10, CI(.02, .17), p=.01), and discomfort (b=.23, CI(.05, .23), p=.002) were each associated with increased self-perceived bias. In contrast, positive affect was associated with increased perceptions of the credibility, objectivity, and validity of the IAT (b=.19 CI(.09, .29), p<.001) and agreement with the test results (b=.25, CI(.07, .43), p=.006), and other-directed negative affect was associated with less favorable assessments of social scientists (b=-.16, CI(-.27, -.06), p=.002).

Main Effects of Experimental Manipulation on Motivated Reasoning

The effect of experimental condition on each indicator of motivated reasoning is represented in Figure 2.7. These results indicate that participants in the bias *(b* = -.18 (*95% CI = -.24, -.12) p* <.001) and accuracy feedback conditions *(b* = -.17 (*95% CI = -.24, -.11) p* <.001), compared to the no feedback condition, were less likely to regard the

IAT as a credible, valid, and objective measure of their implicit attitudes. However, no other significant differences in motivating reasoning were observed across the experimental conditions.

Effects of Individual Differences on Motivated Reasoning

It was expected that **e**xtrinsic egalitarians and highly prejudiced individuals would be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias. Thus, people characterized by either 1) explicitly hostile racial attitudes, 2) resentment towards racial progress, 3) system justification, 4) social dominance orientation, and 5) high levels of extrinsic egalitarianism relative to intrinsic egalitarianism are expected to be motivated to reject evidence of their own implicit racial bias. To evaluate the psychological predictors of the motivated rejection of implicit racial bias, each subscale of FMR-B was regressed separately on each indicator of explicit racial attitudes, sociopolitical orientations, and egalitarian motivations. Figure 2.8 provide a coefficients plot that graphically represent the unstandardized coefficients and 95% CI for each individual difference measure from models estimating its impact on self-perceived implicit racial bias, perceptions of social scientists, perceptions of the IAT, agreement with test results (feedback conditions only) and belief in prejudice and discrimination, respectively.

These analyses suggest that extrinsic (vs. intrinsic) egalitarianism (b=-.18, CI(-.31, -.05), p=.006) was associated with self-perceived implicit bias. Negative attitudes towards blacks was associated with negative perceptions of the credibility and objectivity of the IAT (b=-.13, CI(-.27, .02), p=.096). However, racial resentment (b=-.13, CI(- .21, -.05), p<.01), negative attitudes towards blacks (b=-.16, CI(-.29, -.03), p<.05), SDO (b=-

.31, CI(-.42, -.19), p<.001), and extrinsic (vs. intrinsic; b=.29, CI(.14, .43), p<.001) egalitarian motivations was associated with less favorable assessments of social scientists. Furthermore, extrinsic (vs. intrinsic; b=-.46, CI(-.71, -.21), p<.001) egalitarian motivations as well as implicit racial bias (b=.96, CI(-1.54, -.37), p=.002) were associated with acceptance of the IAT results. Finally, racial resentment (b=-.29, CI(-.36, -.22), p<.001), attitudes towards blacks (b=.10, CI(-.01, .21), p=.077), SDO (b=-.38, CI(-.50, -.27), p<.001), System Justification (b=-.30, CI(-.43, -.18), p<.001), and extrinsic (vs. intrinsic; b=.29, CI(.17, .41), p<.001) egalitarian motivations were associated with decreased belief in the existence of prejudice and discrimination.

Effects of T1 Individual Differences x Message Condition on Motivated Reasoning

Here, I test a series of hypothesized 2-way and 3-way interactions between individual differences and implicit bias feedback. First, the effect of explicit racial attitudes and sociopolitical orientation on the motivated rejection of feedback about one's own implicit racial bias was expected to be stronger among individuals who received implicit bias feedback (vs. control). Second, it was hypothesized that intrinsic (vs. extrinsic) egalitarians with dissociated implicit-explicit racial attitudes would be more accepting of evidence of their own implicit racial bias, which could be indexed by increased self-perceived bias and/or perceptions of the credibility of the IAT and social science in general, but especially in the bias feedback conditions. It is not expected, but possible, however, that implicit-explicit dissociation will predict increased scores on the FMR-B among extrinsic egalitarians. Finally, I expect high prejudiced individuals who are motivated to project an egalitarian self-image primarily for extrinsic reasons or in compliance with socially imposed norms and expectations to be motivated to reject

evidence of their own unconscious racial bias, particularly when provided with feedback about their own implicit bias.

To test these hypotheses, experimental condition was first interacted separately with egalitarian motivations, sociopolitical orientations, and implicit or explicit racial attitudes to predict each subscale of the FMR-B. Next, the interaction between egalitarianism and implicit-explicit dissociation was submitted to OLS regression to predict each subscale of the FMR-B. Finally, the 3-way interaction between implicit-explicit dissociation, egalitarianism, and experimental condition is examined. Only significant models are reported below.

I begin by first reporting significant interactions between experimental condition and sociopolitical orientations or racial attitudes. The interaction between SDO and 1) bias (vs control) feedback (b=.23, CI(.01, .45), p=.042), and 2) bias (vs. accuracy) feedback (b=-.33, CI(-.59, -.07), p=.014) obtained significance for belief in the existence of prejudice and discrimination. The relationship between SDO and belief in prejudice and discrimination, for each experimental condition, is represented in Figures 2.9. At 1 SD above the mean of SDO, the effect of bias feedback (vs. control and accuracy) was not significant (ps>.25; at 1 SD below the mean of SDO the effect of bias feedback (vs. control) was significant (b=-.11, CI(-.22, .004), p=.041). Furthermore, the relationship between SDO and belief in prejudice and discrimination was significant in the control condition (b=-.42, CI(-.58, -.25), p<.001), accuracy feedback condition (b=-.54, CI(-.78, -.29), p<.001), and bias condition (b=-.21, CI(-.38, -.03), p=.022), although inspection of the unstandardized coefficients suggests that the relationship is weakest in the latter.

Thus, bias feedback attenuated the relationship between SDO and the belief in racial prejudice and discrimination.

Next, I examined the interaction between experimental condition and egalitarian motivations. Results indicate that the interaction between egalitarian motivations and 1) bias (vs. control; b=-.65, CI(-.97, -.33), p<.001) and 2) accuracy (vs. control; b=-.45, CI(-.83, -.06), p=.022) obtained significance for perception of the IAT. The relationship between egalitarianism and perception of the IAT, for each experimental condition, are represented in Figures 2.10. Analysis of the relationship between egalitarian motivations and perceptions of the IAT separately for each condition suggests that in the no feedback control condition, intrinsic (vs. extrinsic) egalitarianism was associated with increased perceptions of the IAT as credible and objective (b=.24, CI(-.02, .50), p=.068). However, the direction of this relationship reversed course in the bias feedback condition (b=-.35, CI(-.59, -.10), p=.007), but did not obtain significance in the accuracy feedback (b=-.16, CI(-.50, .17), p=.336).

Similarly, the interaction between egalitarian motivations and 1) bias (vs. control; b=-.22, CI(-.46, .02), p=.074) and 2) accuracy (vs. bias; b=.25, CI(-.06, .55), p=.110) obtained significance for belief in prejudice and discrimination. The relationship between egalitarianism and belief in prejudice and discrimination, for each experimental condition, is represented in Figure 2.11. Analysis for each condition suggests that in the bias feedback control condition, intrinsic (vs. extrinsic) egalitarianism was unrelated to beliefs in the existence of prejudice and discrimination (b=.09, CI(-.10, .27), p=.346). However, intrinsic (vs. extrinsic) egalitarianism was associated with increased belief in

the existence of prejudice and discrimination in the no feedback condition (b=.27, CI(.10, .44), p=.003) and in the accuracy feedback (b=.35, CI(.06, .64), p=.018).

Thus, bias feedback had differential effects on perceptions of the IAT and belief in discrimination as a function of egalitarian motivations. In particular, bias feedback caused people high in intrinsic and low in extrinsic egalitarianism to devalue the credibility and objectivity of the IAT and attenuated the relationship between egalitarian motivations and the belief in prejudice and discrimination. I obtained no evidence of a 2-way or 3-way interaction between implicit-explicit dissociation and either egalitarian motivation or experimental condition to predict Motivated Reasoning.

Affect as a Mediator for Effect of Implicit Bias on Motivated Reasoning

Because 1) implicit bias feedback induced negative affect and undermined positive affect, and because 2) the former was associated with self-perceived implicit bias and the latter was associated with perceptions of the IAT and acceptance of the IAT results, I conducted a series of mediation analyses to better understand the relationship between implicit racial bias feedback, affect, and motivated reasoning. For this analysis, I examine the mediating role of overall positive or overall negative affect for the effect of receiving implicit racial bias feedback (whether accurate or not; vs. control)[4] on each indicator of the FMR-B [5]. Consistent with the results of Pilot 2, I expect negative affect to mediate the relationship between feedback and self-perceived bias. To do so, I ran a

[4] The effect of either implicit bias or accurate implicit bias feedback on negative and positive affect was significantly different from the control condition, but not each other. To increase statistical analyses, I compare all individuals who received biased feedback to participants who received no feedback in the mediation analyses described in this section.

[5] Because only participants in the accurate or implicit bias feedback received any post-IAT feedback, agreement with the tests results is not included in mediation analysis.

bootstrap analysis that tested for the indirect effect of each predictor on the dependent variable, retaining all covariates.

The results of this analysis indicate that negative affect and, to a lesser extent, positive affect were each a critical mediator for the effects of implicit bias feedback on self-perceived bias and perceptions of the IAT, respectively. Specifically, with overall negative affect submitted as a mediator, I observe significant indirect effect of implicit racial bias feedback on self-perceived implicit bias *(b = .04, SE = .02, (95% CI = .01, .08), p*=.026*),* but not perceptions of the IAT *(b = .02, SE = .01, (95% CI = -.01, .05), p*=.25*),* perceptions of social scientists *(b = -.001, SE = .01, (95% CI = -.03, .02), p*=.92*),* or belief in the existence of prejudice and discrimination *(b = .01, SE = .01, (95% CI = -.01, .05), p*=.36*).* Furthermore, with positive affect submitted as a mediator, I did not observe a significant indirect effect of implicit racial bias on any outcome of interest.

Anti-Bias Interventions, Prejudice-regulation, Stereotyping, and Public Policy Attitudes

The next set of analyses examines the relationship between 1) each subscale of the FMR-B, explicit racial attitudes, sociopolitical orientations, and egalitarian motivations, and 2) attitudes towards anti-bias interventions, prejudice-regulation, stereotyping, and public policy attitudes I hypothesized that the motivated rejection of implicit racial bias would undermine prejudice-regulation, as indexed by attitudes towards anti-bias interventions and the motivations to control prejudice reactions, and increase racial stereotyping and opposition to public policy intended to help racial minorities. Each independent variable was regressed separately on each dependent variable using ordinary least squares regression. Figure 2.13 and 2.14 provides a

coefficients plot that graphically represent the unstandardized coefficients and 95% CI for each independent variable obtained from each model.

Consistent with the results of Study 1 and my hypotheses, the results from these analyses clearly indicate that perceptions of social scientists, belief in prejudice and discrimination, sociopolitical orientations, and egalitarian motives are each significant predictor of anti-bias interventions, stereotyping, and public policy attitudes in ways consistent with my theoretical expectations. In particular, high levels of self-perceived implicit bias (b=.22, CI(.04, .40), p=.018), favorable assessments of social scientists (b=.40, CI(.20, .59), p<.001), belief in prejudice and discrimination (b= .98, CI(.84, 1.12), p<.001), favorable assessments of the IAT (b=.36 , CI(.23, .49), p<.001), and intrinsic (vs. extrinsic) egalitarian motivations (b=.60, CI(.43, .78), p<.001) were associated with increased perceived importance and willingness to participate in anti-bias interventions; in contrast, racial resentment (b=-.40, CI(-.51, -.28), p<.001), SDO (b=-.60, CI(-.77, -.42), p<.001), and system justification (b=-.35, CI(-.53, -.16), p<.001), were associated with devaluation and decreased willingness to participate in anti-bias interventions. However, only favorable assessments of social science (b=.14, CI(.02, .25), p=.02), belief in prejudice and discrimination (b=.24, CI(.12, .35), p<.001), favorable perception of the IAT (b=.08, CI(.0001, .16), p=.05), and intrinsic (vs. extrinsic) egalitarianism (b=.10, CI(-.01, .21), p=.077) was associate with increased motivations to control prejudice reactions; in contrast, racial resentment (b=-.19, CI(-.26, -.12), p<.001) and SDO (b=-.22, CI(-.31, -.12), p<.001) were associated with decreased motivations to control prejudice reactions.

A similar pattern emerged for stereotype endorsement and public policy attitudes. More specifically, belief in the existence of prejudice and discrimination (b=-.78, CI(-.94, -.61), p<.001), favorable perceptions of social science (b=-.51, CI(-.69, -.33), p<.001), and intrinsic (vs. extrinsic) egalitarianism (b=-.60, CI(-.75, -.44), p<.001) were each associated with a reduction in stereotyping; in contrast, racial resentment (b=.64, CI(.56, .73), p<.001), SDO (b=.72, CI(.55, 88), p<.001), and system justification (b=.43, CI(.26, .61), p<.001) was associated with increased stereotyping. Similarly, belief in the existence of prejudice and discrimination (b=.54, CI(.42, .66), p<.001), favorable perceptions of social science (b=.37, CI(.24, .50), p<.001), and intrinsic (vs. extrinsic) egalitarianism (b=.41, CI(.29, .53), p<.001) were each associated with increased support for public policy intended to help racial minorities; in contrast, racial resentment (b=-.39, CI(-.46, -.33), p<.001), SDO (b=-.56, CI(-.67, -.46), p<.001), and system justification (b=-.33, CI(-.45, -.21), p<.001) was associated with decreased support for public policy attitudes intended to help racial minorities.

Next, I test the hypothesis that sociopolitical orientations and racial attitudes were activated by the experimental condition to influence stereotyping. In Study 1, exposure to information about implicit racial bias increased the effect of SDO, system justification, and explicit racial attitudes on racial stereotypes. Here, a similar set of analyses is undertaken, this time evaluating whether these constructs moderate the impact of more personalized feedback about one's own implicit racial bias on prejudice-regulation and stereotyping. This analysis failed to obtain any evidence to suggest that implicit bias feedback strengthened the relationship between sociopolitical orientations or explicit racial attitudes and racial stereotyping.

However, I find evidence to suggest that implicit bias feedback moderated the effect of attitudes towards blacks and system justification on prejudice regulation. In particular, the interaction between experimental condition and attitudes towards blacks obtained marginal significance for attitudes towards anti-bias interventions (bias v. control, b=-.38, CI(-.80, .03), p=.069) and significance for the general motivation to control prejudice reactions (bias v. control, b=-.25, CI(-.50, -.01), p=.045; accuracy v. bias, b=-.28, CI(-.54, -.02), p=.034). To better understand this dynamic, I estimated the effect of attitudes towards blacks on general motivations to control prejudice reaction and attitudes towards anti-bias interventions separately in each experimental condition. The relationship between attitudes towards blacks and prejudice-regulation, for each experimental conditions, is represented in Figure 2.15. The results of this analysis suggest that only in the bias condition did attitudes towards blacks predict attitudes towards anti-bias interventions (b=.30, CI(.02, .59), p=.037).

Similarly, the interaction between experimental condition and system justification obtained significance for attitudes towards anti-bias interventions (bias v. control, b=.65, CI(.20, 1.10), p=.005; accuracy v. control, b=-.49, CI(-.97, -.01), p=.047). To better understand this dynamic, I estimated the effect of system justification on attitudes towards anti-bias interventions separately in each experimental condition. The relationship between system justification and prejudice-regulation, for each experimental conditions, are represented in Figure 2.15. The results of this analysis suggest that only in the accuracy (b==-.41, CI(-.78, -.044), p=.028) and bias (b=-.50, CI(-.81, -.19), p=.002) condition did system justification predict attitudes towards anti-bias intervention.

Motivated Reasoning as a Mediator for Effect of Individual Differences on Racial Stereotyping

In Study 1, I demonstrated that perceptions of social scientists and belief in prejudice and discrimination was found to mediate the relationship between 1) sociopolitical attitudes, egalitarian motivations, explicit racial attitudes, and 2) endorsement of negative racial stereotypes[6]. As in Study 1, results from Study 2 demonstrate that perceptions of social scientists and belief in racial prejudice and discrimination each predict stereotype endorsement, and are predicted by a broad range of variables that were found to also associate with stereotyping (including explicit racial attitudes and sociopolitical orientations), suggesting that these constructs are good candidates as mediators. In order to examine whether these dimensions of motivated reasoning mediates the relationship between 1) explicit racial attitudes, sociopolitical orientations, egalitarian motivations, and stereotype threat, and 2) stereotype endorsement, I ran a bootstrap analysis that tested for the indirect effect of each predictor on the dependent variable, retaining all covariates. Figure A provides a conceptual model of the expected pattern of mediation, in which beliefs about prejudice and discrimination and attitudes about social science each serve as a mediator for the relationship between individual differences and prejudice-regulation, stereotyping, and public policy attitudes.

The results of this analysis indicate that belief in racial prejudice and perceptions of social scientists were each a critical mediator for the effects of individual differences on stereotype endorsement. Specifically, with belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects on stereotype endorsement

for system justification *(b = .20, SE = .05, (95% CI = .11, .30), p<*.001*),* social dominance orientation *(b = .21, SE = .04, (95% CI = .13, .30), p<*.001*),* racial resentment *(b = .11, SE = .03, (95% CI =.06, .17), p<*.001*),* and intrinsic (extrinsic) egalitarian motivation *(b = -.17, SE = .05, (95% CI =-.26, -.08), p<*.001*).* Similarly, with perceptions of social scientists submitted as a mediator, I observe significant indirect effects on stereotype endorsement for social dominance orientation *(b = .09, SE = .03, (95% CI = .03, .16), p<*.01*),* racial resentment *(b = .04, SE = .02, (95% CI =.01, .08), p=*.01*),* attitudes toward blacks *(b = .09, SE = .04, (95% CI =.02, .17), p<.05),* intrinsic (vs. extrinsic) egalitarian motivations *(b = -.09, SE = .04, (95% CI =-.16, -.02), p=*.008*),* but not system justification (p>.15).

Because perceptions of social science and belief in the existence of racial prejudice were also important predictors of both indicators of prejudice-regulation, I also examine the extent to which these constructs mediate the relationship between sociopolitical orientations, racial attitudes, or egalitarian motivations on attitudes towards anti-bias interventions and general motivation to control prejudice reactions using the same procedure for mediation analyses described above.

The results of this analysis indicate that belief in racial prejudice and perceptions of social scientists were each a critical mediator for the effects of individual differences on intervention attitudes and general motivations to control prejudice reactions. Specifically, with belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects on attitudes towards anti-bias interventions for system justification *(b = -.29, SE = .07, (95% CI = -.43, -.17), p<*.001*),* social dominance orientation *(b = -.33, SE = .06, (95% CI = -.44, -.22), p<*.001*),* attitudes towards blacks

(b=.10, SE= *.06* (95% CI =-.01, .21), p=.074), racial resentment *(b = -.25, SE = .04, (95% CI =-.33, -.18), p*<.001*),* and intrinsic (vs. extrinsic) egalitarian motivation *(b = .22, SE = .04, (95% CI =.11, .33), p*<.001*).* Similarly, with perceptions of social scientists submitted as a mediator, I observe significant indirect effects on intervention attitudes for social dominance orientation *(b = -.07, SE = .03, (95% CI = -.14, -.01), p*<.05*),* racial resentment *(b = -.04, SE = .02, (95% CI =-.08, -.01), p*=.037*),* attitudes toward blacks *(b = -.07, SE = .03, (95% CI =-.15, -.01), p<.05),* intrinsic (vs. extrinsic) egalitarian motivation *(b = .06, SE = .03, (95% CI =.01, .13), p*=.033*),* but not system justification (p>.15).

For motivations to control prejudice reactions, perceptions of social scientists appears to be a less consistent mediator for the effect of individual differences on the general motivation to control prejudice reactions than for other mediation analyses. With belief in prejudice and discrimination submitted as a mediator, I observe significant indirect effects for system justification *(b = -.07, SE = .02, (95% CI = -.12, -.03), p*<.01*),* social dominance orientation *(b = -.06, SE = .02, (95% CI = -.11, -.01), p*<.001*),* racial resentment *(b = -.04, SE = .02, (95% CI =-.08, .004), p*=.076*),* intrinsic (vs. extrinsic) egalitarian motivation *(b = .06, SE = .02, (95% CI =.02, .10), p*=.004*),* but not attitudes towards blacks (p>.1). Similarly, with perceptions of social scientists submitted as a mediator, I only observed significant indirect effects on general motivations to control prejudiced reactions for intrinsic (vs. extrinsic) egalitarians *(b = .03, SE = .02, (95% CI* =-.001, .06), p=.089), but not attitudes towards blacks (p>.1), SDO (p>.25), system justification (p>.25) or racial resentment (p>.2).

Study 2 Discussion

Study 2 was designed to manipulate the belief that one harbors implicit racial bias. Using a post-IAT feedback paradigm and experimental stimuli designed to be similar in believability and accuracy, no reliable evidence was observed in Study 2 to suggest that feedback suggesting that one is characterized by implicit racial bias (whether accurate or not) impacted self-perceived implicit bias, perceptions of social scientists, or belief in the existence of racial prejudice. Thus, implicit bias feedback did not *directly* increase awareness of the propensity for oneself or others to act with prejudice, nor did it incur backlash against social scientists who study racial prejudice and discrimination. However, consistent with my hypothesis that White Americans are motivated to reject evidence of implicit racial bias, participants in both feedback (compared to no feedback) conditions were significantly less likely to regard the IAT as an objective, valid, or credible measure of their racial attitudes.

In addition to increasing the motivation to devalue the validity of the IAT, implicit bias feedback (whether accurate or not) increased negative affect and reduced positive affect. These affective consequences of implicit bias feedback are critical, as negative affect mediated its relationship to self-perceived implicit bias. Furthermore, the accuracy of the implicit feedback did not alter these effects, suggesting that incongruency between the feedback and "accurate" self-knowledge is insufficient to explain this observed pattern of results. However, the one exception to this pattern is in regards to other-directed negative affect, which was heightened in the accuracy feedback condition relative to the other two conditions. It's possible that individuals who have accurate self-knowledge of their own bias are particularly resentful of others when reminded of this bias. Additional research should explore this possibility in greater detail. Nonetheless, the

overall pattern of results is clear. Implicit bias feedback (whether accurate or not) induced negative affect and promoted the motivated rejection of the source of that information, and, via its affective impact, increased self-awareness of one's own implicit bias. Importantly, these main effects and mediation analyses from Study 2 are consistent with the pattern of results obtained from an independent sample that was recruited for Pilot 2.

I also observed several main effects of individual differences on motivated reasoning. Consistent with the results from Study 1, extrinsic (vs. intrinsic) egalitarians were more likely to perceive their own bias, and, along with people characterized by SDO, System Justification, and negative explicit racial attitudes, were less likely to believe in prejudice and discrimination or hold favorable opinions of social scientists. Furthermore, this same pattern of results extended to perceptions of the IAT; system justification, SDO, and explicit racial bias were each associated with rejection of the IAT as a valid instrument for measuring racial attitudes. Thus, individuals who are motivated to appear unprejudiced for extrinsic (vs. intrinsic) reasons or justify the legitimacy of the status quo or social hierarchy, or who harbor explicitly negative attitudes towards racial minorities, and were particularly motivated to reject feedback regarding their implicit racial bias. These relationships between sociopolitical orientations and racial attitudes and the motivated rejection of implicit bias did not vary across experimental condition, suggesting that implicit bias feedback may not be effective among individuals characterized by such beliefs.

In contrast, the main effect of implicit bias feedback on perceptions of the IAT was observed most strongly among individuals characterized by intrinsic (vs. extrinsic) egalitarianism, suggesting that, consistent with my hypotheses, egalitarian motivations

patterns the effect of implicit bias feedback on motivated reasoning in this context-individuals who have internalized egalitarian social norms into their self-concepts and individuals who were *not* preoccupied with avoiding the appearance of being prejudiced were particularly motivated to devalue the source of implicit bias feedback. Similarly, intrinsic (vs. extrinsic) egalitarianism was less strongly related to belief in prejudice and discrimination in the implicit bias feedback condition relative to the other experimental conditions. While the moderating role of intrinsic egalitarianism on perceptions of the IAT and belief in discrimination is consistent with my expectations (i.e., individuals who truly value egalitarianism are threatened by implicit bias feedback), the direction of effect of in the biased feedback conditions came as a surprise, as I expected intrinsic (vs. extrinsic) egalitarians to be less motivated to reject evidence of implicit bias. It is possible that one feature of impression-management strategies designed to appear unprejudiced is to accept feedback of one's own bias and to recognize patterns of discrimination in society, not to deny it. However, whereas intrinsic (vs. extrinsic) egalitarianism was associated with less self-perceived bias, it was also associated with increased recognition of prejudice discrimination in society, independent of feedback condition. Additional work should seek to replicate this anomalous finding.

Consistent with Study 1, the effect of explicit racial attitudes and sociopolitical orientations on motivated reasoning processes did not vary across experimental condition, suggesting that these individual differences may promote relatively stable attitudinal orientations about racial bias, discrimination, and inequality. However, in Study 1, I found that exposure to information about implicit racial bias increased the effects of explicit racial attitudes, SDO, and System Justification on racial stereotyping.

People who believe in the legitimacy of social hierarchy, perceive the status quo as fair and desirable, and hold explicitly negative attitudes towards blacks were more motivated to attribute the causes of racial inequality in the criminal justice system, employment decisions, and socio-economic status to personal (vs. situational) factors.

In Study 2, I tested the hypothesis that exposure to implicit bias feedback would strengthen the association between these individual differences and stereotyping and prejudice regulation. However, I did not observe an interaction between these constructs and bias feedback condition on stereotyping, suggesting that, unlike exposure to general information about implicit bias, personalized feedback does not activate racial stereotyping as a function of these individual differences. However, I find evidence to suggest that implicit bias feedback moderated the effect of attitudes towards blacks and system justification on prejudice regulation. In particular, individuals with negative attitudes towards blacks held more favorable attitudes towards anti-bias interventions when exposed to implicit bias feedback. In contrast, people motivated to justify the legitimacy of the status quo held less favorable attitudes towards anti-bias interventions.

Finally, while belief in the existence of prejudice and discrimination and perceptions of social sciences were relatively stable orientations unaffected by personalized implicit bias feedback, these beliefs nonetheless mediated the relationship between individual differences and racial stereotyping and prejudice-regulation, which, in turn, was associated with public policy attitudes. Consistent with the results of Study 1, individuals characterized by negative explicit racial attitudes, belief in the legitimacy of the status quo and social hierarchy, or were characterized by extrinsic (vs. intrinsic) egalitarianism were 1) less likely to believe in the existence of prejudice and

discrimination, and held more negative attitudes towards social science, and, consequently, 2) were more likely to engage in racial stereotyping, less motivated to engage in prejudice regulation, and thereby more likely to oppose public policy designed to improve the lives of racial minorities.

Given these downstream consequences of motivated reasoning in this context, it is of critical importance to identify strategies to attenuate defensive responding and increase awareness of implicit racial bias. For this reason, Study 3 was undertaken.

Pilot 3: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback

The experimental paradigm used in Study 3 was first pilot tested on a sample drawn from the target population. Pilot 3 employed a single independent variable design (Implicit Racial Bias Feedback: No Intervention vs. Self-Affirmation vs. Collective Bias). The primary goal of the pilot for Study 3 was to independently validate the pre-feedback intervention paradigms. Furthermore, Pilot 3 will provide preliminary evidence of the extent to which the interventions are able to reduce the affective and cognitive consequences of implicit racial bias feedback observed in Study 2, and that each intervention successfully manipulate targeted beliefs and attitudes. In general, I do not expect any differences between conditions in terms of the perceived believability or accuracy of the information presented in the experiment. However, manipulation check items specific to each intervention are designed to differentiate among participants assigned to each condition.

Method

Participants

Participants were 357 White U.S. citizens recruited from Amazon MTurk (61% females, 39% males; mean age = 37.44, *SD* = 12.77). Most participants were modestly affluent (49% report a family income greater than 50K) and educated (45.7% have earned at least a Bachelor's degree). With this sample size, to detect mean level differences between the experimental and control group, I estimated that I had 46% power to detect a Cohen's *d* of 0.2, 99% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit White U.S. citizens and would compensate participants $0.50 for their time. The name of the study is intended to increase the expectation that one's beliefs and attitudes about other people would be directly measured.

The procedure for Pilot 3 is an abridged version of the procedure for Study 3. In particular, I do not include a post-IAT feedback control group. Since my focus in Pilot 3 is evaluating the interventions, *all* participants in Pilot 3 will receive post-IAT feedback indicating that they harbor implicit racial bias. Prior to receiving feedback, however, participants in Pilot 3 will be randomly assigned to either the self-affirmation intervention, "collective bias" intervention, or control group.

Participants assigned to the self-affirmation condition first completed the general motivation to control prejudice responding scale (e.g., Dunton & Fazio, 1997). This measure was used in Study 1 and 2 as a dependent variable, but has been found in other research to effectively affirm participants' egalitarian values and reduce the threat of

being stereotyped as a racist (e.g., Frantz et al., 2004). Next, participants rank-ordered a list of characteristics and values, and then wrote 1-3 paragraphs about the 1st-ranked value and why it is important to them (adapted from Schmeichel & Martens, 2005). Finally, participants then completed a series of manipulation check items.

Participants assigned to the "collective bias" intervention received the following instructions, prior to post-IAT feedback:

It is important to understand that this test does NOT *guarantee* that you are racially biased, nor does it mean that you have discriminated against racial minorities in the past. While unconscious racial bias is extremely common and is quite normal, it is also something that, once you are aware of it, you are able to control. For example, this test has been administered to a very large sample of the people in countless different studies. The results from these studies indicate that the overwhelming majority of people harbor unconscious racial bias- even among people who strongly support racial equality and value racial tolerance. However, people who were made aware of their implicit bias were also better able to control it and minimize its influence on their judgment and behavior.

Social and behavioral scientists agree that unconscious preferences for some racial groups are a normal, basic feature of human cognition, and it has reliably been observed across most cultures and historical periods. In fact, one study determined that even social scientists who study racial discrimination commonly harbor unconscious racial prejudice. Most psychologists believe that unconscious beliefs, like the beliefs measured by this test, reflect the information available in the social environment and not some deep-rooted bigotry or hatred towards people in society. In this sense, unconscious racial bias is a basic feature of human cognition. It is a common and normal consequence of living in modern times, but it also something that people are able to control, once they become aware that it is influencing their thoughts and behavior.

Next, participants in the collective bias intervention condition completed a series of reading comprehension questions, which are designed to reinforce the substance of the intervention and assess accurate understanding of its content.

After receiving treatment from the intervention condition, participants will then receive the post-IAT feedback. Participants in the control intervention condition

responded to the manipulation checks for both interventions conditions and then proceeded straight to the post-IAT feedback.

Finally, participants completed measures of affect, general manipulation checks, motivated reasoning, the Stroop task (described in Study 2), and demographic measures, before being fully debriefed.

Measures

See Tables 3.1 and 3.2 for means, SD, and correlations between all measures used in analyses for Pilot 3. The measures were identical to that used in Pilot 2. However, Pilot 3 also employed unique measures to assess the effectiveness of the manipulation. These new measures are described below. As before, all measures are recoded to run from 0-1 (except the scores on the Stroop task).

Self-Affirmation Manipulation Checks. 2 sets of items were used to validate the self-affirmation manipulation. 7 items were used to measure the extent to which the essay writing task made participants think about aspects of the self that are important, positive, likable or desirable. Higher values represent more positive thoughts about the self (M= .58, SD=.20 , Cronbach's Alpha=.88). 6 items were used to measure the extent to which participants felt positively about themselves, included clever (vs. foolish), adequate (vs. inadequate), good (vs. bad), important (vs. unimportant), inferior (vs. superior), and unattractive (vs. attractive). Higher values represent more positive feelings about the self (M=.55, SD=.17 , Cronbach's Alpha=.89).

Collective Bias Manipulation Checks. Participants answered 4 true or false items designed to measure comprehension of the information contained in the collective bias prompt. Such items include, "According to psychological scientists, unconscious

prejudice is extremely common in the American population", "Most psychological scientists agree that unconscious racial prejudice in basic feature of human cognition", "Prior research indicates that even social scientists who study race relations harbor unconscious racial prejudice", and "Unconscious beliefs reflect the information in the social environment, and not some deep-rooted bigotry or hatred towards racial minorities". Higher values represent a larger number of correct responses (M=.89, SD=.23, Cronbach's alpha=.61).

Results

Two dummy-coded variables were created to represent condition assignment, with the no intervention condition as the reference group. For analyses comparing differences between the two intervention conditions, an additional set of dummy-coded variables were created such that the self-affirmation condition served as the reference group, although this was not the primary focus of my analysis. For all of the models used in analyses reported below, both dummy-coded variables are included along with covariates.

Manipulation Checks

Participants' judgments of the believability, accuracy, and deceptiveness of the information contained in the experiment did not differ across the three conditions ($p>.2$). However, participants in the self-affirmation (vs. no intervention) condition were more likely to report positive thoughts *(b* = .17 (*95% CI = .11, .21) p* <.001) and beliefs *(b* = .08 (*95% CI = .13, .04) p* <.001) about the self, providing strong evidence for the validity of the self-affirmation manipulation. Similarly, participants in the collective bias (vs. no intervention) condition were more likely to accurately report that unconscious

racial prejudice is common and fundamental to human cognition *(b* = .11 (*95% CI = .06, .17) p* <.001). Thus, I obtained strong evidence to suggest that the interventions successfully and uniquely manipulated targeted beliefs and affect.

Main Effects of Experimental Manipulation on Affect and Motivated Reasoning

Next, I examined the effect of the experimental manipulation on affect and motivated reasoning. Results did not indicate any significant differences in self-reported affect or motivated reasoning across intervention conditions. However, participants in the collective bias intervention (vs. control) were significantly more likely to believe in the existence of prejudice and discrimination (b=.05, CI(.00, .10), p=.053) and held marginally more favorable attitudes towards social scientists (b=.04, CI(-.01, .09), p=.086).

Relationship Between Affect and Motivated Reasoning

Figure 3.1 represents a coefficient plot for the estimated relationship of each affect dimension and each indicator of motivated reasoning. These results indicate that, on a sample of participants who all received implicit bias feedback, positive affect and each indicator of negative affect predicted self-perceived implicit bias- participants who reported positive affect were *less* likely to perceive their own bias, whereas participants who experience negative affect were *more* likely to report their own bias. Negative affect, but no other affective measure were associated with more favorable perceptions of the IAT and with agreement of the test results. Positive affect undermined belief in prejudice and discrimination, whereas all indicators of negative affect (except other-directed negative affect), was associated with increased belief in prejudice and discrimination. Finally, positive affect and other-directed negative affect was associated with less

positive evaluations of social scientists. Thus, negative affect, which can be induced by bias feedback, appears able to increase self-perceived bias, belief in the existence of prejudice and discrimination in society, acceptance of the test results, and perceptions of the IAT as credible.

Effect of Motivated Reasoning on Cognitive Depletion

The difference in response latency for incongruent and congruent trials was regressed on the dummy-coded variables for condition, with the implicit bias –no intervention condition as the reference group. These results indicate that cognitive depletion (i.e., slower response latency to incongruent vs. congruent Stroop trials) was reduced in the collective bias condition *(b* = -.50.96 (*95% CI = -99.77, -2.16*), *p* =.041).

Discussion

Consistent with the results of Study 2, post-feedback negative affect was associated with increased self-perceived implicit bias, more favorable perceptions of the IAT, and increased belief in discrimination and prejudice; in contrast, positive affect was associated with more favorable assessments of social science but decreased belief in prejudice and discrimination. However, self-reported post-IAT affect did not vary across experimental condition. Nevertheless, participants in the collective bias (vs. control) condition were more likely to believe in the existence of prejudice and discrimination and held more favorable attitudes towards social scientists, suggesting this intervention may help reduce defensive responding and the motivated rejection of implicit racial bias.

More importantly for the purposes of Pilot 3, the experimental stimuli are appropriate for use in Study 3. Perceptions of the believability, accuracy, and deceptiveness of the information contained in the experiment did not vary across

condition. Furthermore, participants in the self-affirmation (vs. no intervention) condition were more likely to report more favorable thoughts and beliefs about the self; participants in the collective bias (vs. no intervention) condition were more likely to accurately report that unconscious racial prejudice is common and fundamental to human cognition, consistent with the consensus opinion of psychological scientists. For these reasons, I have sufficient confidence in the validity of the experimental paradigm used in Study 3.

Study 3a: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback

In Study 3a, I independently manipulate "self-affirmation" and "collective bias" interventions to evaluate the relative success of each in reducing motivated reasoning processes and its effects found in Study 1 and Study 2. This study employs a 1 (Post-IAT No Feedback Control) + 3 (Post-IAT Bias Feedback, No Intervention; Post-IAT Bias Feedback, Self-Affirmation Intervention; Post-IAT Bias Feedback, Collective Bias Intervention) design, in which a subset of participants are randomly assigned to an intervention prior to the receipt of implicit bias feedback. In particular, I expect to replicate the results obtained in Study 2 by comparing the results for participants in the no feedback condition with the bias feedback, no intervention condition[7]. However, I also expect that the impact of implicit bias feedback on affect and motivated reasoning processes, as well as its interactions with individual differences for these outcomes, will be attenuated in the intervention condition relative to the bias feedback, no intervention condition. I do not have any expectations regarding the success of both interventions

[7] The correlation and mediation analyses in Study 3 are consistent with what was observed for Study 1 and Study 2. For this reason, I do not report correlation analyses for Study 3, and focus only on the main effects of experimental condition or its interaction with individual differences.

relative to each other. Accordingly, I do not compare participants in the two intervention conditions together, and focus only on differences between the 1) intervention conditions or no feedback conditions, and 2) bias feedback, no intervention condition.

Unfortunately, there was a coding error in the T2 survey for Study 3a, and several of the critical measures of motivated reasoning were excluded. For this reason, my analyses here focus only on the affective consequences of implicit bias feedback, the success of the interventions in attenuating these processes and the interaction between interventions and individual differences. In Study 3b, I address the cognitive consequences of implicit bias feedback and the role of both the pre-feedback interventions and individual differences in this dynamic.

Method

Participants

Participants were 802 White U.S. citizens recruited from Amazon MTurk (63% females, 37% males; mean age = 35.60, *SD* = 12.87). Most participants were modestly affluent (45% report a family income greater than 50K) and educated (43.9% have earned at least a Bachelor's degree). Of these individuals, 407 were retained at T2 (51%)[8]. With this sample size, to detect mean level differences between the experimental and control group, I estimated that I had 41% power to detect a Cohen's *d* of 0.2, 97% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

[8] This high rate of attrition is very concerning, as it can produce misleading and unreliable estimates (Zhou & Fishbach, 2016). This is a major limitation for interpreting the pattern of results in Study 3a, which has a substantial higher attrition rate than Study 1 and 2.

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit white U.S. citizens for a two-part study, and would compensate participants $1.50 total for their time (T1=$0.25, T2=$1.25).

This study used a 2-wave panel design. Pre-manipulation measures were assessed at Time 1 (T1). Participants first viewed a consent form, and complete the pre-manipulation battery. Up to one week later, participants were contacted to complete the second part of the study. At Time 2 (T2), participants were randomly assigned to the no feedback, implicit bias feedback-no intervention, self-affirmation interventions, or collective bias intervention (described in Study 2 and Pilot 3).

Measures

See Table 3.3 for means, SD, and correlations between all measures used in analyses for Study 3a. The T1 and T2 measures were identical to that used in Study 2.

Results

For comparisons between the implicit bias feedback and no feedback condition, which are intended to replicate the results of Study 2, a dummy-coded variable was created to represent condition assignment, with the no feedback control condition as the reference group. For analyses comparing differences between the implicit bias feedback-no intervention condition with each of the two intervention condition, an additional set of dummy-coded variables was created such that the bias feedback-no intervention condition serves as the reference group. Interaction terms were constructed by taking the product between the dummy-coded condition variable and rescaled continuous predictor variables. For moderation analyses, I primarily examine the effect of the moderating variable separately for each condition when simple slopes analyses fail to clarify the

nature of the interaction. All statistical models include age, gender, income, and trust in social science as covariates.

Main Effects of Experimental Manipulation on Affect

Next, I examined the effect of the experimental manipulation on affect. Marginal means for each measure of affect, separate for each condition, are represented in Figure 3.2 or 3.3. Results replicate the findings from Study 2, and indicate that participants in the bias feedback condition, compared to the no feedback conditions, reported significantly lower levels of positive affect, and significantly higher levels of negative affect overall, discomfort, depression, and self-directed negative affect, but did not report any difference in other-directed affect. No significant differences in affect were observed between participants in the 3 bias feedback conditions for positive affect, depressed affect, discomfort, or other-directed affect (ps>.15). However, participants in the self-affirmation (vs. bias no intervention) condition experienced marginally less overall negative affect (b=-.06, CI(-.13, .01), p=.086) and self-directed negative affect (b=-.08, CI(-.17, .003), p=.06), suggesting that self-affirmation may minimize self-directed negative affect due to implicit bias feedback.

Effects of T1 Individual Differences x Message Condition on Affect and Self-Perceived Bias

Here, I test a series of hypothesized 2-way and 3-way interactions between individual differences and implicit bias feedback, with the intention being to replicate the results of Study 2 and to evaluate the extent to which the interventions attenuate these outcomes. Given the missing measures of motivated reasoning, I can only evaluate the effect on affect and self perceived bias for Study 3a- tests of the remaining hypotheses

are undertaken in Study 3b. First, it was hypothesized that feelings of guilt and shame among low prejudiced, intrinsic (vs. extrinsic) egalitarians would attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation. A similar set of analyses was conducted with SDO, System Justification, and implicit and explicit racial attitudes to explore their role as moderators for the effect of feedback on affect. Furthermore, it was expected that **e**xtrinsic (vs. intrinsic) egalitarians and highly prejudiced individuals would be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias. Thus, people characterized by either 1) explicitly hostile racial attitudes, 2) resentment towards racial progress, 3) system justification, 4) social dominance orientation, 5) high levels of extrinsic egalitarianism relative to intrinsic egalitarianism are expected to be motivated to reject evidence of their own implicit racial bias. Finally, I also test the hypothesis that sociopolitical orientations and racial attitudes were activated by the experimental condition to influence stereotyping.

In Study 2, I did not find any evidence to suggest that the impact of implicit bias feedback on affect varies as a function of experimental condition or its interaction with racial attitudes or egalitarian motivation. In contrast, the results of Study 3a suggest that the interaction between the bias feedback-no intervention condition (vs. no feedback-no intervention) and attitudes towards blacks and egalitarian motivations obtained significance for *each* indicator of negative affect. In particular, bias feedback-no intervention (vs. no feedback or intervention) condition interacted with attitudes towards blacks (b=-.67, CI(-1.09, -.25), p=.002) and intrinsic (vs. extrinsic) egalitarianism (b=.66, CI(.32, .99), p<.001) to predict negative affect. The estimated effect of each individual

difference on overall negative affect, separately for the two conditions, is presented in Figures 3.4 and 3.5. Additional analyses suggest that the effect of attitudes towards blacks on negative affect marginally differed in the bias feedback-no intervention (b=-.26, CI(-.56, .04), p=.092) vs. control (b=.30, CI(-.02, .62), p=.067) conditions; similarly, the effect of intrinsic (vs. extrinsic) on negative affect differed in the bias feedback-no intervention (b=.24, CI(.03, .45), p=.023) vs. control (b=-.40, CI(-.69, -.10), p=.009). Together, these results suggest that implicit bias marginally increased negative affect, especially among individuals with negative racial attitudes and with intrinsic (vs. extrinsic) egalitarian motivations.

Because I had originally hypothesized a *discrete* affective response following implicit bias feedback as a function of egalitarian motivations---such that intrinsic egalitarians would experience self-directed negative affect and extrinsic egalitarians would experience other-directed negative affect—I examined the interaction on these dependent variables more closely (see Figure 3.6). However, the pattern of results do not support the hypothesized discrete emotional responses experiences by intrinsic vs. extrinsic egalitarians in the implicit bias feedback condition. Indeed, bias feedback induced negative affect among intrinsic (vs. extrinsic) egalitarians, whether that is self-directed or other-directed.

Furthermore, in Study 2, I found that implicit bias feedback undermined positive and induced negative affect among individuals low in SDO. The results of Study 3a support replicate this finding. In particular, for SDO interacted with bias-no intervention (vs. control) condition to predict positive affect (b=.43, CI(-.01, .87), p=.054), overall negative affect (b=-.60, CI(-1.03, -.16), p=.007), and each dimension of negative affect.

These results are represented in Figure 3.7. Additional analyses reveal that, for positive affect, the impact of SDO differed in the implicit bias-no intervention (b=.13, CI(-.43, .17), p=.40) and control condition (b=.38, CI(.05, .70), p=.024); similarly, the impact of SDO on overall negative affect differed in the implicit bias-no intervention feedback (b=.42, CI(.12, .73), p=.007) and control condition (b=.01, CI(-.31, .33), p=.95). Together, these results suggest that implicit bias feedback reduce positive affect and increased negative affect as a function of SDO. Critically, the effect of SDO, attitudes towards blacks, or egalitarian motivations on affect did not differ among participants who received implicit bias feedback as a function of the interventions.

Finally, in Study 2, I also find evidence to suggest that implicit bias feedback moderated the effect of attitudes towards blacks and system justification on prejudice regulation. In Study 3a, I find some evidence that interventions patterned the impact of system justification on self-perceived bias, stereotyping, and policy attitudes. In particular, the interaction between system justification and the self-affirmation intervention (vs. bias feedback control) condition was significant for self-perceived bias (b=.31, CI(.04, .58), p=.023), stereotyping (b=.29, CI(.05, .52), p=.016), and policy attitudes (b=-.38, CI(-.64, -.11), p=.006). Similarly, the interaction between system justification and the collective bias intervention (vs. bias feedback control) condition obtained significance for stereotyping (b=.28, CI(-.05, .61), p=.003), and policy attitudes (b=-.34, CI(-.60, .09), p=.009). Figure 3.8 represents the estimated effect of system justification on each of these outcomes, separately for all 3 bias feedback conditions. However, I did not observe an interaction between these constructs and implicit bias feedback no intervention (vs. no feedback, no interventions), and therefore do not have

consistent evidence to support what was observed in Study 2, suggesting that much of the analyses reported below could be capitalizing on chance.

Additional analyses reveal that the effect of system justification on self-perceived implicit bias did not obtain significance when evaluated separately within each of the implicit bias feedback conditions. However, simply slope analyses reveal that at 1 SD above the mean of system justification, the effect of the self-affirmation condition (b=-.09, CI(-.16, -.01), p=.027), but not the collective bias intervention (b=-.03, CI(-.10, .03), p>.3), compared to bias-no feedback condition, on self-perceived bias was significant; at 1 SD below the mean of system justification, compared to the bias-no feedback condition, the impact of self-affirmation (b=-.20, CI(-.37, -.04), p=.017), but not collective bias (b=-.10, CI(-.25, .06), p=.225), on self-perceived bias was significant. Furthermore, the effect of system justification on stereotyping was stronger in the self-affirmation (b=.54, CI(.27, .81), p<.001) and collective bias interventions (b=.66, CI(.39, .92), p<.001), than in the bias feedback-no intervention condition (b=.31, CI(.05, .56), p=.018), Similarly, the effect of system justification on opposition to public policies was significant in the self-affirmation (b=-.61, CI(-.79, -.42), p<.001) and collective bias intervention (b=-.44, CI(-.59, -.30), p<.001), but not in the bias feedback-no intervention condition (b=-.14, CI(-.35, .09), p=.232). Together, these results suggest that self-affirmation intervention (vs. bias feedback no intervention) reduced self-perceived implicit bias among individuals both high and low in system justification, although the effect was largest among individuals low in System Justification. Furthermore, both interventions increased the effect of system justification on racial stereotyping and opposition to public policies

designed to help racial minorities, raising the possibility that these interventions may be counter-productive.

Study 3a Discussion

Study 3a was designed to replicate the results of Study 2 and evaluate the effectiveness of two interventions –self-affirmation and collective bias—in reducing motivated reasoning and defensive responding to implicit bias feedback and increasing awareness. The results replicate the finding observed in Study 2 that implicit bias feedback increased negative and decreases positive affect, but had no impact on self-perceived bias. Furthermore, the observed impact of implicit bias feedback on negative affect was stronger among individuals with negative racial attitudes and intrinsic (vs. extrinsic) egalitarians motivations. While this latter finding is consistent with my hypotheses, I did not observe this pattern of results in Study 2, suggesting that additional efforts to test this hypothesis are needed. Furthermore, I expected and found in both Study 2 and Study 3a that implicit bias feedback was particularly likely to undermine positive and induce negative affect among individuals low in SDO.

Neither affect nor self-perceived bias varied as a function of participation in the collective bias (vs. bias feedback no intervention) condition, although I did observe a marginally significant reduction of negative affect among participants in the self-affirmation condition. Interestingly, the self-affirmation intervention reduced self-perceived implicit bias for individuals at 1 SD above and below the mean of system justification, although this effect was larger for individuals low in system justification. That self-affirmation reduced both negative affect and self-perceived bias (although only at 1 SD above/below mean of system justification) is not surprising, given that the results

of Study 2 indicated that negative affect mediated the relationship between implicit bias feedback and self-perceived bias. Finally, both interventions increased the deleterious effects of system justification on stereotyping and opposition to policies intended to help minorities, suggesting that these interventions may backfire for individuals motivated to justify the legitimacy of the status quo.

However, due to a coding error that led to 3 (i.e., perceptions of the IAT, perceptions of social scientists, and belief in discrimination) of the 4 dimensions of motivated reasoning to not be included in the survey, I could only evaluate the effects of implicit bias feedback and interventions, and its interactions with individual differences, on affect and self-perceived bias. For this reason, I conducted a condensed version of Study 3a to explore the main and interactive effect of individual differences, bias feedback, and the interventions on affect and all dimensions of motivated reasoning, which will allow me to replicate and extend the findings from Study 2 and Study 3a.

Study 3b: Interventions to Reduce Motivated Rejection of Implicit Bias Feedback

In Study 3b, as for Study 3a, I independently manipulate "self-affirmation" and "collective bias" interventions to evaluate the relative success of each in reducing motivated reasoning processes and its effects found in Study 1 and Study 2. This study employs a 1 (Post-IAT No Feedback Control) + 3 (Post-IAT Bias Feedback, No Intervention; Post-IAT Bias Feedback, Self-Affirmation Intervention; Post-IAT Bias Feedback, Collective Bias Intervention) design, in which a subset of participants are randomly assigned to an intervention prior to the receipt of implicit bias feedback. I plan to test the same hypotheses as for Study 3a, but to also explore the effects of implicit bias feedback and the interventions, as well as their interactions with individual differences,

on *all* indicators of motivated reasoning- not just self-perceived implicit bias. Measures of prejudice-regulation, stereotyping, and public policy attitudes were not included in Study 3b.

Method

Participants

Participants were 356 White U.S. citizens recruited from Amazon MTurk (63% females, 37% males; mean age = 35.38, *SD* = 13.57). Most participants were modestly affluent (48.9% report a family income greater than 50K) and educated (46.6% have earned at least a Bachelor's degree). With this sample size, to detect mean level differences between the experimental and control group, I estimated that I had 38% power to detect a Cohen's *d* of 0.2, 96% power to detect a Cohen's *d* of 0.5, and 99% power to detect a Cohen's *d* of 0.8.

Procedure

Participants were recruited for a study of "Attitudes About People". The study advertised that it was primarily looking to recruit white U.S. citizens for a study, and would compensate participants $.50 total for their time.

This study used a single survey design that is a condensed version of Study 3a. Racial resentment, system justification, social dominance orientation, egalitarian motivations, and trust in social science were assessed after consent and prior to the IAT and random assignment to condition. After the IAT and experimental treatment (described in Pilot 3 and Study 3a), participants completed measures of affect and each measure of motivated reasoning.

Measures

See Table 3.4 for means, SD, and correlations between all measures used in analyses for Study 3b. The measures were identical to that used in Study 2.

Results

Analyses for Study 3b were conducted the same way as for Study 3a.

Main Effects of Experimental Manipulation on Affect

First, I examined the effect of the experimental manipulation on affect. Marginal means for each measure of affect, separate for each condition, are represented in Figure 3.9. Results replicate the findings from Study 2 and Study 3a, and indicate that participants in the bias feedback condition, compared to the no feedback conditions, reported significantly lower levels of positive affect, and significantly higher levels of negative affect overall, discomfort, depression, and self-directed negative affect, but did not report any difference in other-directed affect. However, unlike Study 3a, participants in the self-affirmation condition (b=.08, CI(-.003, .16), p=.061) and collective bias intervention (.09, CI(-.002, .28), p=.054), compared to the bias feedback-no control condition, experienced higher levels of positive affect. No other differences in affect were observed across the bias feedback condition.

Main Effects of Experimental Manipulation on Motivated Reasoning

In Study 3a, I could only test the main effect of experimental condition on one of the four primary indicators of motivated reasoning, and did not observe any effect on self-perceived bias. Consistent with the results of Study 2 and 3a, I do not observe any difference in self-perceived bias for participants in the bias feedback- no intervention condition compared to the no bias feedback control (b=-.004, CI(-.06, .05), p=.89). Consistent with the results of Study 2, I also find that in the bias feedback-no intervention

(vs. control), participants held more negative attitudes towards the IAT (b=-.21, CI(-.27, -.15), p<.001), providing strong support for the hypothesis that individuals would be motivated to discredit the IAT upon receipt of implicit bias feedback. However, unlike Study 2, I also observe significant effects on perceptions of social scientists (b=-.04, CI(-.09, .00), p=.048), and belief in prejudice and discrimination (b=-.09, CI(-.14, -.04), p=.001), suggesting that implicit bias not only increased the motivated rejection of the IAT, but may also lead to backfire effects against social scientist and undermine perceptions of prejudice and discrimination in society.

More critically for the purposes Study 3b, I obtain robust evidence to suggest that the collective bias intervention (but not self-affirmation) successfully reduced the motivated rejection of implicit bias. In particular, compared to the bias feedback-no intervention condition, collective bias intervention increased self-perceived bias (b=.07, CI(.01, .13), p=.015), more favorable perceptions of the IAT (b=.14, CI(.08, .21), p<.001), acceptance of the test results (b=.14, CI(.05, .23), p=.002), and belief in prejudice and discrimination (b=.07, CI(.02, .11), p=.007), but had no impact on perceptions of social scientists (b=.02, CI(-.02, .07), p=.283). No differences between the self-affirmation and bias feedback-no intervention condition were observed for any of these indicators. Thus, the collective bias intervention successfully attenuated the motivated rejection of implicit racial bias. Marginal means for each measure of Motivated Reasoning, separate for each implicit bias feedback condition, are represented in Figure 3.10 and 3.11.

Effects of T1 Individual Differences x Message Condition on Affect

I had hypothesized that feelings of guilt and shame among low prejudiced, intrinsic (vs. extrinsic) egalitarians would attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation. However, I've obtained inconsistent evidence from Study 2 and Study 3a suggesting that intrinsic (vs. extrinsic) egalitarians were more likely to experience negative affect upon receipt of implicit bias feedback, and no evidence to suggest that self- or other-directed negative affect was uniquely related to intrinsic or extrinsic egalitarianism, respectively. A similar set of analyses was conducted with SDO, system justification, and explicit racial attitudes to explore their role as moderators for the effect of feedback on affect. It is possible that individuals who believe in the legitimacy of social hierarchy or the status quo, or who hold negative explicit racial attitudes, will be less likely to show increased negative affect upon receipt of implicit bias feedback. In Study 2 and Study 3a, I obtained consistent evidence to suggest that implicit bias feedback reduced positive and induced negative affect among individuals low in SDO, and I seek to replicate this finding in Study 3b.

Consistent with the results of Study 2 and 3a, I find that SDO interacts with the implicit bias-no intervention (vs. control) condition to predict positive (b=.72, CI(.26, 1.18), p=.002), overall negative affect (b=-.50, CI(-.93, -.07), p=.022), self-directed negative affect (b=-.71, CI(-1.20, -.22), p=.005), and discomfort (b=-.48, CI(-.95, .00), p=.050). The estimated effect of SDO on both positive and negative affect failed to obtain significance when examined separately for each condition. However, simple slope analyses suggest that, for individuals 1 SD above the mean of SDO, implicit bias-no feedback (vs) control decrease positive affect (b=-.28, CI(-.36, -.20), p<.001) and increase negative affect (b=.10, CI(.03, .16), p=.004); at 1 SD below the mean of SDO,

the implicit bias-no feedback (vs) control also decreased positive affect (b=-.60, CI(-.82, -.38), p<.001) and increased negative affect (b=.32, CI(.12, .52), p=.002). While the effect was observed at both 1 SD above/below the mean of SDO, inspection of the coefficients suggests that the effect was larger among individuals low in SDO, thereby confirming the results of Study 2 and Study 3a.

Next, I examined the extent to which the interventions differentially influenced the impact of implicit bias feedback on affect as a function of SDO. I find that SDO interacted with the collective bias (vs. bias feedback-no intervention) condition to predict positive (b=-.67, b=-1.16, -.17), p=.009), but not negative affect. At 1 SD above the mean of SDO, the collective bias intervention (vs. bias feedback-no intervention) increased positive affect (b=.08, CI(-.002, .17), p=.056); at 1 SD below the mean of SDO, this effect also obtained significance (b=.38, CI(.13, .62), p=.003). Together, these results demonstrate the implicit bias feedback can increase negative affect and reduce positive affect among individuals both high and low in SDO, although the effect is larger for the latter group. Further, the collective bias intervention mitigated the impact of implicit bias feedback on positive affect, although this too was stronger among individuals low in SDO. More importantly, the collective bias intervention did not attenuate the impact of implicit bias feedback on negative affect across levels of SDO. Because negative affect mediates the relationship between feedback and self-awareness, it is critical to note that this intervention did not attenuate negative affect across levels of SDO. No other interaction obtained significance.

Effects of T1 Individual Differences x Message Condition on Motivated Reasoning

It was expected that **e**xtrinsic (vs. intrinsic) egalitarians and highly prejudiced individuals would be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias. Thus, people characterized by either 1) explicitly hostile racial attitudes, 2) resentment towards racial progress, 3) system justification, 4) social dominance orientation, 5) high levels of extrinsic egalitarianism relative to intrinsic egalitarianism are expected to be motivated to reject evidence of their own implicit racial bias. In Study 2, I obtained evidence to suggest that bias feedback caused people high in intrinsic and low in extrinsic egalitarianism to devalue the credibility and objectivity of the IAT, and attenuated the relationship between 1) both system justification and egalitarian motivations and 2) the belief in prejudice and discrimination. In Study 3a, I also found evidence to suggest that the self-affirmation manipulation reduced self-perceived implicit bias among individuals low in system justification.

Unlike Study 2, I do not find any evidence to suggest that implicit bias feedback led intrinsic egalitarians to devalue the IAT or to change their belief in the existence of prejudice and discrimination. In Study 1, Study 2, and Study 3a, I find consistent evidence to suggest that extrinsic (vs. intrinsic) egalitarianism was associated with decreased belief in prejudice and discrimination. For that reason, the extent to which the interventions can attenuate the tendency for extrinsic egalitarians to underestimate prejudice and discrimination in society is of critical importance, even if these beliefs are not activated by implicit bias feedback. Accordingly, I find that intrinsic (vs. extrinsic) egalitarianism interacted with the collective bias (vs. bias feedback-no intervention control) intervention to predict belief in prejudice and discrimination (b=-.35, CI(-.60, -

.09), p=.008). The relationship between egalitarian motivations and belief in prejudice and discrimination across implicit bias feedback conditions is represented in Figure 3.12. Additional analyses reveal that the effect of intrinsic (vs. extrinsic) egalitarianism obtained significance in the bias feedback-no intervention condition (b=.21, CI(.01, .41), p=.036). However, in the collective bias intervention, the relationship between egalitarian motivations is in the opposite direction, although it was not significant (b=-.14, CI(-.31, .03), p=.12). Thus, the collective bias intervention appears able to attenuate the tendency for extrinsic egalitarianism to underestimate the existence of prejudice and discrimination in society.

I also find that the interaction between SDO and bias feedback (vs. control) failed to obtain significance for belief in prejudice and discrimination (b=-.24, CI(-.55, .07), p=.125), like in Study 3a. Nevertheless, in Study 1, 2, and 3a, SDO was associated with decreased belief in prejudice and discrimination, regardless of experimental condition. Furthermore, the interaction between collective bias (vs. bias feedback-no intervention) was marginal significant for belief in discrimination (b=.29, CI(-.01, .59), p=.06). Accordingly, I estimated the effect of SDO on belief in prejudice separately across implicit bias feedback condition. This analyses indicates that the effect of SDO on belief in prejudice was larger in implicit bias-no intervention condition (b=-.43, CI(-.66, -.19), p=.001) than in the collective bias condition (b=-.20, CI(-.36, -.03) p=.022) or control condition (b=-.25, CI(-.46, -.04), p=.021). These results suggest that bias feedback activates the effect of SDO on disbelief in prejudice and discrimination, but the collective bias intervention attenuated this effect, as it did for extrinsic egalitarian motivations. The

relationship between SDO and belief in prejudice and discrimination across implicit bias feedback conditions is represented in Figure 3.13.

Furthermore, the interaction between implicit bias feedback (vs. control) and SDO obtained significance for perceptions of the IAT (b=.42, CI(.09, .75), p=.013). The interaction between collective bias (vs. implicit bias no intervention) also obtained significance (b=-.49, CI(-.87, -.11), p=.012). Additional analyses indicate that the effect of SDO on perceptions of the IAT only obtained significance in the implicit bias feedback-no intervention condition (b=.49, CI(.19, .79), p=.002), suggesting that bias feedback activates the effect of SDO on rejection of the IAT, but the collective bias intervention attenuated this effect. The relationship between SDO and perceptions of the IAT across implicit bias feedback conditions is represented in Figure 3.14. While many of these observed effects are interesting and consistent with my expectations, I did not observe these patterns in Study 2, raising the distinct possibility that these are anomalous findings. Additional research and/or meta-analytical technique for collapsing across samples may be necessary to determine the extent to which these findings are consistent across samples.

Study 3b Discussion

Study 3b was designed to replicate the results of Study 2 and to extend Study 3a to evaluate the effectiveness of two interventions –self-affirmation and collective bias— in reducing the affective and cognitive consequences of implicit bias feedback. The results replicate the finding observed in Study 2, Pilot 3, and Study 3a that implicit bias feedback increased negative and decreases positive affect (especially among individuals low in SDO), promotes negative reactions to the IAT, but has no impact on self-perceived

bias. However, unlike Study 2, I also find that implicit bias feedback undermined perceptions of social scientists and belief in prejudice and discrimination, suggesting that implicit bias feedback may incur additional backfire effects, although the failure to observe this pattern in Study 2 suggests that this may be a relatively small or unreliable effect.

More importantly, I obtain robust evidence to suggest that the collective bias intervention successfully reduced defensive responding following receipt of implicit bias feedback. Furthermore, the collective bias intervention did not attenuate the impact of implicit bias feedback on negative affect, which is a reliable mediator for the relationship between feedback and self-awareness. In contrast, I find some evidence to suggest that self-affirmation prior to implicit bias feedback can increase positive affect and bolster against negative affect, but no evidence to suggest that this intervention impacted the cognitive consequences of implicit bias feedback. Altogether, the results do not support the value of self-affirmation as an intervention to reduce defensive responding in this context. In contrast, that the collective bias interventions reduced defensive responding, increased positive affect, but did not protect individuals from the negative affect induced by implicit bias feedback, strongly supports the utility and efficacy of this strategy**.** Furthermore, one major concern with the collective bias intervention was the risk of communicating the normative acceptability of implicit forms of racial bias and undermining individuals' motivation to adjust for its influence on their judgment and behavior, leading to ironic effects by undermining prejudice-regulation motivation and the perceived value of anti-prejudice interventions (Duguid & Thomas-Hunt, 2014). Fortunately, in Study 3a, I did not observe any evidence of increased stereotyping or a

reduction in prejudice-regulation among participants in the collective bias condition, suggesting that this intervention is unlikely to increase belief in the acceptability of prejudicial beliefs or discriminatory behavior.

Future research should manipulate specific features of the collective bias intervention to better understand the psychological processes by which learning that implicit bias is both common and fundamental to human cognition can reduce defensive responding to evidence and feedback about implicit bias. For example, in Study 1, exposure to the scientific information about implicit bias did not necessarily lead to increased awareness or acknowledgement of prejudice in society. Instead, such information either had no impact on self-awareness, or it led to more negative attitudes towards social sciences or increased stereotyping among individuals with negative racial attitudes and a motivation to justify the status quo and existing social hierarchy. However, when similar information was paired with more personalized implicit bias feedback, defensive responding was reduced, suggesting that both personalized feedback *and* information about the science of implicit bias may be necessary, if not sufficient, conditions for reducing motivating reasoning in this context.

The collective bias intervention used in Study 3 also incorporated several additional components beyond personalized implicit bias feedback and information about the science of implicit bias and racial prejudice. For example, the collective bias condition also reassured participants that they were able to *control* the expression of implicit bias and its consequences for their behavior, and that prejudiced responding was *not* inevitable, so long as one is aware of the potential for such bias and is motivated to reduce it. Furthermore, participants were informed that, according to social scientists,

implicit bias is a reflection of the information available in the social environment, and not some "deep-rooted bigotry or hatred", thereby reducing the perception that one is uniquely to blame for harboring unconscious beliefs that are inconsistent with prevailing egalitarian norms and values. Thus, the collective bias interventions also "normalized" the existence of implicit bias (without communicating its acceptability within society), appealed directly to egalitarian values (by assuming a reduction in bias was desirable), reduced negative social comparisons or distinctiveness from one's social group (i.e., Americans), and, maybe most critically, informed participants that such bias is within their control, if they are able and willing to recognize it.

It's possible that the success of the collective bias intervention depended on the presence of each of these features, although this remains an empirical question to be addressed by future research. In addition to personalized feedback and information about the science of bias, however, I suspect that communicating the "controllability" of such bias is also of critical importance. Prior research on defensive responding to persuasive information, especially information that threatens one' self-concept (i.e., fear-appeals; Howell & Shepperd, 2012), emphasizes the importance of instilling a sense of efficacy regarding one's ability to successfully change their attitudes and behavior to avoid negative or undesirable consequences (Witte & Allen, 2000). In the context of implicit bias, providing individuals with strategies to overcome biases has also been found to be among the most successful approaches (Lai et al., 2014). Furthermore, individuals who are able and willing to control the expression of bias can be successful in doing so (Sassenberg & Moskowitz, 2005; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). Together, this literature strongly points to the importance of the perceived controllability

of implicit bias as a mechanism for reduced defensive responding and increased prejudice-regulation, including greater acceptance of one's own implicit bias.

However, independent research also suggests that the implicit activation of egalitarian goals reduces the automatic activation of stereotypes (Moskowitz, Salomon, & Taylor, 2000), suggesting that egalitarian appeals may also be an effective anti-bias strategy, although the evidence for this is mixed (Lai et al., 2014). Future research must manipulate each feature of the collective bias intervention orthogonally to explore whether or not these components operate best in conjunction, or if the success of such an intervention depends, instead, on the extent to which participants feel able to control their bias, are motivated to act in ways consistent with egalitarian values and social norms, or feel less personally threatened by awareness of their bias given how common it is in the population. This line of work would not only advance our understanding of ways to reduce defensive responding and increase the success of anti-bias interventions, but would also illuminate the underlying psychological processes associated with the effective regulation of implicit bias.

General Discussion

Three studies investigated the psychological factors implicated in the motivated rejection of implicit racial bias evidence and feedback, and its downstream consequences for the motivation to control prejudicial responding, perceived effectiveness of and willingness to participate in anti-prejudice interventions, endorsement of racial stereotypes, and attitudes about the criminal justice system and public policies intended to help racial minorities. I also tested several strategies to reduce motivated reasoning processes that are engaged when White Americans encounter credible evidence

documenting the existence of implicit bias or when they obtain feedback indicating that, contrary to their egalitarian beliefs and intentions, they harbor unconscious forms of racial bias.

Below, I summarize the key findings and evidence across 8 experiments in which I systematically manipulated exposure to scientific information about racial discrimination and unconscious bias, personalized feedback about one's own implicit bias, and pre-feedback interventions designed to reduced defensive responding and increase awareness. I also examine the assumption that self-awareness is necessary for the effectiveness of anti-bias interventions, consider the implications of these findings for existing models of prejudice-regulation and the motivational basis of racial attitudes, explore the application of my results to political and legal contexts, and discuss future directions for this program of research.

How do White Americans respond to information and feedback about implicit racial bias?

Anti-bias interventions commonly share the assumption that increasing awareness of racial prejudice and implicit bias is an effective strategy for reducing its impact on judgment and behavior. Furthermore, existing models of prejudice-regulation (e.g., Czopp, Monteith & Mark, 2006; Monteith, 1993) also suggest that *only* when individuals become aware of their propensity for prejudice responding are they able to control its expression. Thus, a major research question addressed by this program of research is whether or not exposing White Americans to scientific information or direct personalized feedback about the existence of racial prejudice and unconscious biases increases

awareness or induces backfire effects in the form of motivated reasoning processes, and whether individual differences can explain this phenomena.

Effect of exposure to scientific information on awareness

Results from three experiments (i.e. Pilot 1a, Pilot 1b, Study 1) indicates that exposing White Americans to scientific information about implicit racial bias *can* increase self-perceived implicit bias, but may also promote more unfavorable opinions of social scientists and increase the effect of socio-political and racial attitudes on endorsement of racial stereotypes. However, the effect of the manipulation on self-perceived bias was inconsistent across the pilots and main experiments for Study 1—compared to information about health and exercise (but not information about the automaticity of psychological functioning), scientific information about implicit bias can increase self-awareness. Belief in prejudice and discrimination appears to be least affected by this information, suggesting that these beliefs may be relatively more stable and resistant to new information than self-perceived implicit bias or attitudes towards social science.

However, the effect of explicit racial attitudes, sociopolitical orientations, and egalitarian motivations on motivated reasoning processes did *not* vary across experimental condition. This pattern of results indicates that the experimental paradigm may have been ineffective in promoting the perception of discrimination, whether or not if these individual differences promote a relatively stable attitudinal orientation about racial bias, discrimination, and inequality (i.e., Adams, Tormala & O'Brien, 2006). Future research should identify more effective means of manipulating beliefs about discrimination to better understand if these attitudes are endogenous to intractable

individual differences and are therefore relatively fixed orientations that are resistant to change. Or, if these attitudes are subject to change under conditions not identified by the current research.

Furthermore, defensive responding or the lack of attitudinal change following exposure to scientific information may have emboldened many White Americans who are motivated to justify, rationalize, or altogether ignore the reality of racial discrimination in society. For example, individuals characterized by a belief in the legitimacy of the status quo or in existing social hierarchies, or who held negative explicit racial attitudes, were more likely to engage in racial stereotyping across all conditions, but especially upon receipt of scientific information about implicit bias. These findings highlight the potential backfire effects associated with this paradigm. Consistent with prior work on resistance to persuasion and meta-cognitive factors (e.g., Tormala, Clarkson, & Petty, 2006), it's possible that increased stereotyping as a function of the interaction between these individual differences and information about implicit bias constitutes a psychological mechanism by which existing beliefs about prejudice and discrimination become entrenched. More specifically, future research should explore the extent and conditions under which the presentation of scientific information on implicit bias increases counter-argumentation and favorable resistance appraisals that reinforce one's confidence in the validity of pre-existing beliefs.

Together, the results from Study 1 provide mixed evidence for the utility of awareness raising interventions focused *exclusively* on presenting the scientific information on implicit bias, and instead highlight the potential for such information to induce motivated reasoning and increase racial stereotyping.

Effect of personalized feedback of implicit bias on awareness

Across the pilots and main experiments for Study 2 and 3, personalized implicit bias feedback consistently increased negative affect, reduced positive affect, and increased the devaluation of the IAT. The devaluation of the IAT upon receipt of implicit bias feedback was particularly common among individuals characterized by intrinsic (vs. extrinsic) egalitarianism, even though these individuals were also more likely to engage in prejudice-regulation and prejudice-reduction. I discuss this pattern of results in greater detail in subsequent sections, but this finding supports my contention that even individuals who may genuinely be committed to egalitarian principles may nonetheless be motivated to devalue sources of information and feedback that call into question the extent to which one has acted in ways consistent with prevailing social norms and personal standards for conduct.

I obtain some evidence to suggest that the motivated rejection of implicit bias feedback spilled over into negative attitudes about social scientists, but this finding was inconsistent across samples. Regardless, I obtain no evidence to suggest that implicit bias feedback increased self-perceived bias or belief in the existence of discrimination and prejudice. Together, these results suggest that implicit bias feedback is threatening to White Americans, and activates the motivation to deny the existence and consequences of implicit racial bias. Consistent with this observation, I find that individuals who were particularly motivated to devalue the IAT and social scientists, unwilling or unable to perceive their own bias, or who denied the existence of prejudice and discrimination were significantly more cognitively depleted in this context. Furthermore, this overall pattern of results did not differ between participants who were randomly assigned to receive

implicit bias feedback compared to individuals who received such feedback only if their performance on the IAT indicated bias. Thus, these findings cannot be attributed to the inconsistency between individuals "accurate" self-knowledge of their own implicit attitudes and the feedback provided to them.

While the personalized implicit bias feedback appears to have induced backfire effects without promoting increased awareness, it important to note that negative affect was a critical mediator for the relationship between implicit bias feedback and self-perceived bias. When implicit bias feedback induced negative affect, White Americans became more aware of their own bias. Thus, future research should more closely explore the kinds of individual differences and other factors that might predispose individuals towards negative affect upon learning of their own implicit bias. I return to this issue more below when discussing the implications of these findings for existing models of prejudice-regulation.

In addition to obtaining consistent evidence for the affective and cognitive consequences of the motivated rejection of implicit bias feedback, I also obtain robust evidence for the success of an intervention that reduced defensive responding without attenuating negative affect. In particular, some participants were informed, prior to the receipt of implicit bias feedback, that unconscious bias is common and fundamental to human cognition, and that one is able to control the expression of prejudice if they are able and willing to recognize the influence of implicit bias. This "collective bias" intervention successfully reduced defensive responding among White Americans, such that participants in this condition (vs. implicit bias feedback-no intervention condition) were more willing and able to perceive their own bias, held more favorable attitudes

towards the IAT, and were more inclined to perceive discrimination and prejudice in society. Furthermore, relative to the other two implicit bias feedback conditions, the collective bias intervention reduced cognitive depletion.

While several additional questions remain about the precise psychological mechanism by which this intervention had its impact (see discussion section of Study 3b), these findings nonetheless provide strong evidence for the utility of this strategy for reducing defensive responding. In contrast, I did not find evidence supporting the use of self-affirmation intervention as a means of reducing defensive responding. Affirming the integrity of the self prior to the receipt of feedback did not lead to increased awareness nor more favorable attitudes towards the IAT or social science. Instead, self-affirmation increased positive affect and, to some extent, reduced negative affect. Because negative affect was a reliable mediator for the impact of implicit bias feedback on awareness, the effects of self-affirmation are undesirable. In later sections, I consider the implications of these findings for understanding the motivational basis of belief in and awareness of racial bias and discrimination. Nonetheless, the results of Study 3 strongly support the value of pairing personalized feedback about implicit bias with the information contained in the collective bias intervention.

Does the motivated rejection of implicit racial bias challenge existing models of prejudice-regulation?

Given that implicit bias information and feedback activated motivated reasoning without increasing self-awareness, and the central importance of self-awareness for anti-bias interventions and models of prejudice-regulation, it's critical to evaluate the evidence for the hypothesized effect of awareness for bias reduction. Thus, an additional

research question to which the current research may also speak is whether or not increased self-awareness or increased awareness of discrimination in society directly leads to increased prejudice-reduction or efforts to engage in prejudice-regulation. Here, I focus first on the extent to which exposure to evidence of implicit bias or feedback about one's bias can produce *change* in these constructs and, second, the predictive utility of variability in awareness for prejudice-regulation, stereotyping, and policy attitudes. Finally, I consider the extent to which this body of evidence is consistent with, or not, existing models of prejudice-regulation, with particular focus on the hypothesized role of egalitarian motivations for patterning the effects of exposure to implicit bias information and feedback.

Do information or feedback paradigms increase awareness, and is awareness necessary for prejudice-regulation or prejudice-reduction?

Neither exposure to scientific information about implicit bias nor personalized feedback about one's own bias (when not paired with the collective bias intervention) was sufficient to reliably increase belief in prejudice and discrimination or self-perceived bias. This constitutes strong evidence that White Americans are motivated to deny the existence of implicit racial bias and discrimination. While I only find mixed evidence to suggest that scientific information can increase self-perceived bias, that this construct is not necessarily a fixed attitudinal orientation is an important finding that substantiates the utility of anti-bias interventions that focus on increasing self-awareness of the propensity for prejudice-related discrepancies, at least, to the extent that increased awareness leads to attitudinal and behavioral change (see below). Indeed, results across two independent samples recruited for pilot1a and pilot 1b both indicate that self-perceived implicit racial

biases are malleable and capable of changing in relation to credible information. In contrast, across all samples obtained for this program of research, beliefs about prejudice and discrimination in society were relatively unchanged in the face of credible information about implicit racial bias and its linkages to discrimination in the general population.

Thus, information or feedback paradigms designed to increase self-awareness may not effectively lead to meaningful changes in these beliefs. However, when implicit bias feedback induced negative affect, as I reliably observed across multiple samples, it also increased self-perceived bias, suggesting that the indirect effects of feedback paradigms can lead to increased awareness. Also, when implicit bias feedback was paired with the collective bias intervention, I observed direct effects on both indicators of self-awareness. Together, these findings suggest that personalized feedback, not general scientific information about bias, can directly (when preceded by information contained in the collective bias intervention) or indirectly (via induced negative affect) increase awareness of racial bias and discrimination.

Importantly, the belief in the existence of racial prejudice and self-perceived implicit racial bias predicted an increase in prejudice regulation, regardless of condition, although the former was a more reliable predictor of these outcomes, since it was only associated with attitudes towards anti-bias intervention not the general motivation to control prejudice reactions. Additionally, belief in racial prejudice and, to a lesser extent, perceptions of social science each mediated the relationship between 1) explicit racial attitudes, sociopolitical orientations, and egalitarian motivations and 2) prejudice-regulation and endorsement of racial stereotypes. White Americans who hold explicitly

negative attitudes towards blacks, are motivated to justify the legitimacy of social hierarchy or the status quo, or who are extrinsically (vs. intrinsically) egalitarian were less willing to believe in racial prejudice and held less favorable attitudes towards social science, which increased prejudice-regulation and racial stereotyping and, in turn, bolstered opposition to public policies intended to help racial minorities. This finding is consistent with prior work that has considered the motivational factors associated with belief in the existence of racial discrimination and prejudice (e.g., Adams, Tormala & O'Brien, 2006).

Self-perceived implicit bias did not mediate these relationships, nor did it predict the motivation to control prejudice-reactions, racial stereotyping, or public policy attitudes. Thus, belief in prejudice and discrimination, but not self-perceived bias, mediated the relationship between individual differences and prejudice-regulation and racial stereotyping, the latter of which was a reliable predictor of public policy attitudes. Nevertheless, this overall pattern demonstrates that the motivated acceptance or rejection of implicit bias evidence and feedback is an important determinant of prejudice-regulation, stereotyping, and public policy attitudes.

Causes and Consequences of Self-Perceived Bias

In a limited sense, that increased awareness of one's implicit racial bias did not directly translate into increased motivations to control prejudice reactions, a reduction in racial stereotyping, or support for policies designed to help racial minorities, challenges the general assumption that self-awareness is a necessary condition for effective prejudice-regulation or prejudice-reduction, or that targeting these beliefs should be the focus of anti-bias interventions. However, self-awareness of prejudice-related

discrepancies was multi-determined. In particular, across all samples, extrinsic (vs. intrinsic) egalitarians were more willing and able to perceive their own bias, but *less* willing to engage in prejudice-regulation, reject racial stereotypes, or support public policies intended to help racial minorities. A similar pattern emerged for individuals characterized by negative explicit racial attitudes and a belief in the legitimacy of social hierarchy, although these findings were less consistent across samples as for the role of egalitarian motivations.

Self-perceived implicit bias may therefor reflect different motivations and may have different consequences for different kinds of people. For example, people who hold negative racial attitudes, support social hierarchy, and are anxious about appearing racist may be more willing and able to recognize that they harbor implicit bias, but these characteristics appear to have no direct impact on the motivation to control prejudice responding. People who are intrinsically motivated by egalitarian ideals may be less willing to recognize their own implicit bias, but are nonetheless more motivated to control their prejudice responding, perhaps in part because intrinsic (vs. extrinsic) egalitarian motivations covary with more positive attitudes towards blacks, less racial resentment, and increased belief in prejudice and discrimination, each important predictors of such motivations.

Thus, as a function of stable individual differences, self-perceived implicit bias may emerge in diverse form among individuals who either 1) support social hierarchy and explicitly negative attitudes towards blacks, and are therefore unlikely to adjust for a racial bias that they may regard to be an accurate and fair representation of the social world, 2) are not intrinsically committed to egalitarian ideals and therefore less willing

and able to regulate their own prejudice, 3) are merely anxious about the prospect of appearing racist, but unmotivated to change their behavior to avoid this attribution, or 4) are extrinsically committed to egalitarian ideals, but potentially resentful toward these social norms**,** and therefore less motivated to regulate their prejudiced reactions. The willingness to report or perceive one's own bias appears to be common among the kinds of individuals inclined to believe that such bias is a veridical representation of the social world- an accurate perception of members of specific social groups that is not in need of change, whether at the individual-level or in terms of policies intended for remediation. Because self-perceived implicit bias reflects different motivations captured by distinct individual differences with distinct consequences for attitudes and behavior, it may not directly translate to increased motivations to regulate prejudice, nor, as was observed, to decreased stereotyping or more favorable attitudes towards racially liberal public policy.

The differential role of egalitarian motivations and sociopolitical orientations in this dynamic is of critical theoretical importance. The findings that extrinsic (vs. intrinsic) egalitarian motivations are a more reliable predictor of self-perceived bias than racial attitudes and sociopolitical orientations raises the possibility that individuals who are pre-occupied with concerns about *appearing* racist, not individuals who are motivated to justify and rationalize existing racial inequality, who are able and willing to report a belief in their own bias. If the low predictive utility of self-perceived bias for prejudice-reduction and prejudice-regulation is due to its shared relationship with extrinsic egalitarianism, and not racial attitudes or sociopolitical orientations, it's possible that self-awareness of one's own bias is an impression-management strategy undertaken by individuals motivated to appear unprejudiced, not an expression of sincere belief among

individuals inclined to believe that racial bias is an accurate perception, or natural and desirable outcome, of our social world. That extrinsic (vs. intrinsic) egalitarianism was directly associated with increased self-perceived bias but reduced perceptions of prejudice and discrimination in society provides some early support for this proposition. Additional research is needed to better understand the extent and conditions under which (e.g., direct personal feedback about bias vs. general information about the existence of implicit bias) self-perceived bias reflect motivations to appear unprejudiced versus the intergroup attitudes of individuals motivated to rationalize and justify prejudicial beliefs and racial inequalities as normal, inevitable, and desirable (and therefore not in need of change).

Do these findings support or challenge existing models of prejudice-regulation?

Importantly, this complexity surrounding the causes and consequences of self-awareness of one's propensity for implicit prejudice-related discrepancies is to some extent consistent with existing work on the determinants of prejudice-regulation (e.g., Monteith, 1993), which emphasizes the importance of the internalization of egalitarian social norms as a standard for personal conduct. This work shows that people who have internalized egalitarian standards for their personal conduct become more, not less, motivated to regulate prejudice. White Americans' who are motivated to *project* an egalitarian self-image for extrinsic reasons (e.g., to avoid the attribution of racist intent) or primarily to comply with prevailing egalitarian social norms commonly respond with resentment and denial to information suggesting that they have acted in ways inconsistent with egalitarian social norms (Monteith & Voils, 1998). Similar to this existing body of work, I find consistent evidence that intrinsic (vs. extrinsic) egalitarians and people with

less resentment towards blacks appear more motivated to regulate their own prejudice. In contrast, White Americans who are merely anxious about the prospect of appearing racist or extrinsically motivated egalitarians appear unmotivated to regulate their prejudiced reactions, perhaps because the latter are resentful of racial progress or regard their racial attitudes as veridical representations of the social world.

Consistent with prior research on the impact of self-awareness of the propensity for prejudicial responding (i.e., prejudice-related discrepancies) on affect, it was hypothesized that feelings of guilt and shame among low prejudiced, intrinsic (vs. extrinsic) egalitarians following implicit bias feedback would attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation, particularly among individuals with dissociated implicit-explicit racial attitudes. Similarly, it was possible but not expected that, among extrinsic (vs. intrinsic) egalitarians, experience of other-directed negative affect would increase the motivated rejection of implicit bias following such feedback.

I obtained limited support for these hypotheses. People low in SDO experienced decreased positive affect and increased negative affect when provided with feedback about their own bias. However, inconsistent with my expectations, neither intrinsic nor extrinsic egalitarians were more inclined to experience other- or self-directed negative affect. Thus, for extrinsic egalitarians, resentment towards others was not heightened following exposure to implicit bias feedback, nor did other-directed affect play any role in mediating or predicting motivating reasoning or other downstream consequences of interest. Further, intrinsic egalitarians did not experience increased self-directed negative affect upon receipt of implicit bias feedback, nor was this affective response uniquely

consequential for downstream outcomes of interest. The effect of implicit bias feedback on affect did not vary across condition as a function of egalitarian motivations, system justification, or racial attitudes. Further, the 2-way and 3-way interaction between implicit bias feedback, egalitarian motivations, and implicit-explicit dissociations failed to obtain significance.

Together, this pattern of results does not support the hypothesis that individuals characterized by intrinsic (vs. extrinsic) egalitarian motivations experience self-directed or other-directed negative affect when encountering implicit bias feedback, nor did this effect emerge as a function of implicit-explicit dissociation. However, people who were not inclined to justify the legitimacy of social hierarchy experienced heightened levels of negative affect and reduced positive affect when provided with feedback about their own implicit bias. The failure to observe a unique role of self or other-directed affect as a function of egalitarian motivations and implicit bias feedback is inconsistent with other research (e.g., Monteith & Mark, 2009). However, that overall negative affect mediated the relationship between implicit bias feedback and increased awareness of prejudice-related discrepancies (i.e. self-perceived implicit bias) is nonetheless consistent with existing model of prejudice-regulation, which feature prominently the role of self-awareness for the motivation to act in ways consistent with egalitarian norms.

More generally, this pattern of evidence supports the contention that all White Americans', including self-identified egalitarians, are threatened by and avoidant of information that challenges their unprejudiced self-image (e.g., Frantz et al., 2004; Gaertner & Dovidio, 1986, 2000; Spencer et al., 1998). There are several reasons for why even intrinsic egalitarians are discomfitted and therefore motivated

to *reject* direct evidence of their own implicit bias, instead of responding to this information with a motivation to regulate prejudice and reduce stereotyping (e.g., Monteith, 1993). For example, because implicit attitudes are activated automatically and without conscious awareness, its influence on social judgment is difficult to control or change (e.g., Gregg, Seibt, & Banaji, 2006), let alone recognize. People may perceive a lack of self-efficacy to control the expression of unconscious bias or attenuate its impact on judgment and behavior. A lack of self-efficacy has been found to lead to defensive posturing and behavioral avoidance in response to other kinds of self-threatening information designed to encourage behavioral change (e.g., health domain; Witte & Allen, 2000). This may, in part, explain why the collective bias intervention was successful in reducing defensive responding, as I suggested above.

Further, unlike traditional forms of racial prejudice, contemporary forms of subtle, unconscious racial bias violate lay peoples' intuitions and beliefs about how racial prejudice is expressed and by whom **(**Sommers & Norton, 2006). The existence of one's implicit racial bias is likely to be inconsistent with consciously accessible egalitarian beliefs and attitudes. This kind of self-knowledge may be a more subjectively relevant standard of comparison for evaluating the appropriateness of one's attitudes or behavior than external feedback. Consequently, evidence of one's implicit racial bias may not activate a perceived discrepancy between personal standards and behavior in a way that is sufficient to motivate prejudice-regulation. Indeed, self-reported prejudice-related discrepancies do not appear to be related to implicit measures of racial bias (Monteith, Voils & Ashburn-Nardo, 2001). Thus, for intrinsic egalitarians that truly believe that they are unprejudiced and who are genuinely committed to acting in unprejudiced ways,

feedback suggesting otherwise may rightly be viewed with skepticism. This skepticism may be sufficient to attenuate perceptions of prejudice-related discrepancies and may even lead to the devaluation of the source of such feedback, as I found in Study 2. Nonetheless, these individuals appear able and willing to engage in prejudice-regulation, even if the personalized feedback provided in this context does not necessarily prompt the kind of prejudice-related discrepancies examined in other research. Future research should explore the conditions under which alternative forms of feedback may be sufficient to activate prejudice-related discrepancies among intrinsic egalitarians that will *increase* pre-existing motivations to attenuate expressions of prejudice. I discuss this in greater detail below.

Direct Implications for Political and Legal Contexts

The motivated rejection of implicit racial bias evidence and feedback increased support for racial stereotypes, and, consequently, increased opposition to public policies intended to help racial minorities (e.g., criminal justice reform; social welfare; affirmative action). In this way, my thesis has direct implications for political attitudes. However, identifying obstacles to self-awareness of implicit racial bias, and strategies to overcome it, also have implications for some forms of political communication. For example, subtle racial cues are commonly embedded in political messages (e.g., Huber & Lapinsky, 2006; Mendelberg, 2001), which can contribute to principled opposition to policies designed to help racial minorities (e.g., Gilens, 2001; Hurwitz & Peffley, 2005). Central to the effectiveness of these cues is a lack of awareness of its racial content and its influence on one's thinking. Thus, White Americans who are unable or unwilling to recognize their own implicit racial bias may be particularly vulnerable to racial cues embedded in

political communication. Future research should explore the extent to which personalized feedback about bias, when paired with a "collective bias" intervention, can minimize the influence of racial cues embedded in political communication.

Finally, my thesis research also has potential implications for legal contexts. A clear example of this is the effects of motivated reasoning for the endorsement of racial stereotypes, which can increase support for more punitive criminal justice policies that disproportionately target minority communities (Alexander, 2010; Peffley & Hurwitz, 2010). Racial inequality within the criminal justice system can undermine the public's trust, perceived legitimacy, and compliance with law enforcement, especially among racial minorities (Peffley & Hurwitz, 2010). Identifying strategies to increase White Americans' awareness of implicit racial bias, and legal actors' motivation to regulate prejudice responding, could improve the public's support for criminal justice reform designed to reduce racial disparities within the legal system, or otherwise bolster the perceived legitimacy of the legal process and its institutions (Tyler, 2006). Recently, the Department of Justice announced the implementation of an anti-bias intervention targeting unconscious biases among law enforcement personnel, suggesting that anti-bias interventions will feature prominently in future criminal justice reform efforts. The results of Study 3 indicate that personalized feedback and a "collective bias" intervention may be sufficient to increase awareness and reduce defensive responding. However, future research must replicate these findings in applied contexts among actual legal actors before any concrete recommendations about the likely success of awareness raising interventions can be proffered.

The current work also has implications for legal actors' determination of racial discrimination. Importantly, when determining if discrimination is present in a particular case, legal decision-makers commonly anchor on their commonsense notion of hostile, traditional forms of prejudice that presume behavior is directed by a high degree of self-awareness, deliberation, and intention (Crosby & Dovidio, 2008; Krieger, 2004; Krieger & Fiske, 2006). From this perspective, unfair treatment is viewed as the product of a conscious intention to treat racial minorities unfairly (Banks, Eberhardt, & Ross, 2008). However, this lay theory of behavior and racial discrimination is too narrow a conception of racial discrimination. The existence of implicit racial bias fundamentally challenges these conventional understanding of racial discrimination as conscious and intentional (Kang & Banaji, 2006; Krieger, 2008). Furthermore, intent to discriminate is not necessary for a legally actionable claim of disparate impact in public policy and hiring practices (Bartlett, 2009; *Texas Department of Housing and Community Affairs vs. Inclusive Communities Project*, 2015).

When legal (and other institutional) actors rely upon a lay theory of prejudice that assumes that discrimination is necessarily intentional, they may fail to recognize or ameliorate unintentional, unconscious forms of prejudice (Banks, Eberhardt, & Ross, 2008). As I have argued throughout this paper, such an oversight would be misguided and consequential for individuals, organizations, and institutions committed to antidiscrimination principles and practices (e.g., Jost et al., 2009). However, evidence of implicit bias is commonly introduced into the court room as a "social framework analysis", which provides decision-makers with a context, informed by empirical evidence, for deciding factual issues, such as the presence of racial discrimination (Fiske

& Borgida, 2008; Monahan, Walker, & Mitchell, 2008). Motivated rejection of evidence and feedback of implicit racial bias may impede legal acceptance and recognition of unintentional discrimination. Strategies that allow for the effective communication of evidence of implicit racial bias, without engendering resentment or skepticism, would improve legal actors' ability to recognize and remedy patterns of racial discrimination and disparate treatment that do not conform to prevailing lay theories of prejudice. Future research should more directly explore the extent to which personalized feedback and a "collective bias" intervention can improve legal actors, including jurors, judges, and attorneys, recognition and determination of unintentional discrimination.

Future Directions

The results of these experiments provide the foundation for examining the motivated rejection of evidence of and feedback about implicit bias towards additional disadvantaged target groups, including LGBT members, Women, Muslims, and Hispanics, to name a few. The motivated rejection of implicit bias evidence and feedback towards different target groups, other than racial minorities, may have different kinds of downstream consequences not considered in the present work. For example, the motivated rejection of implicit bias towards women may undermine support for gender diversity in the workplace, equal pay legislation, and family-workplace policies (e.g. Miller, 2014). Future research should explore the psychological processes implicated in the motivated rejection of implicit bias towards other groups (and its consequences) to extend and enrich the current framework.

Similarly, the current research only focused on the experience and psychology of the perceiver or recipient of implicit bias evidence and feedback. Future research that

examines how targets of implicit bias react to perceivers' motivated rejection of implicit bias may also help to identify additional downstream consequences not considered in this work. For example, when targets of prejudice observe dominant groups deny implicit bias evidence, they may infer that society is tolerant of some forms of discrimination. It is possible that this inference could undermine the perceived legitimacy of political and legal institutions, belief in the malleability of group characteristics and power dynamics (e.g., Rydell, Hugenberg, Ray, & Mackie, 2007), and reduce political interest and community engagement among racial minorities. Alternatively, the inference that society is tolerant of discrimination against one's social group may increase the perception and experience of injustice, which could motivate disadvantaged groups to engage in collective forms of action that directly challenges existing social inequality (e.g., Dixon et al., 2012). In this way, White Americans' ignorance and denial of more subtle manifestations of bias could ironically motivate disadvantaged groups to organize against the status quo. Future research should explore these possibilities.

Furthermore, the findings from these studies should not be constrained to contexts in which individuals receive feedback on their performance of *measures* of racial bias. Additional work can explore whether these findings can generalize to others contexts in which individuals are confronted with feedback about their own racial prejudice that is inconsistent with their egalitarian self-image, or may not otherwise conform to prevailing cultural and cognitive representations of what constitutes racial prejudice. These alternative forms include direct interpersonal feedback or confrontation from targets of discrimination or advocates and educators of equality (e.g., Czopp, Monteith, & Mark, 2006; Kaiser & Miller, 2003). The type or source of feedback is only one variable not

considered in this work that could potentially moderate defensive responding and the motivated rejection of bias. It is also possible that White Americans', particularly extrinsic egalitarians, will be even more motivated to reject implicit racial bias feedback when they are publically accountable for complying with egalitarian social norms. To test these hypotheses, future research could manipulate the source of the feedback and the belief that such information may be shared with a public audience.

It is critical to note that the motivated rejection of implicit racial bias evidence and feedback, as well as the reduction in defensive responding among individuals assigned to the collective bias intervention, was only observed at a single time point. It remains for future research to determine the extent to which, individuals for whom, and conditions under which these effects persist across contexts and time, or not. For example, because beliefs about implicit bias and discrimination are strongly associated with stable individual differences, it is possible that any short-term fluctuations in these beliefs will dissipate over time. Repeated exposure to scientific evidence and personalized feedback about implicit bias would therefore presumably be necessary to observe stable long-term change in these beliefs. However, perhaps the persistence of these effects matter less that then their short-term implications for increased compliance with and internalization of anti-bias interventions and strategies for prejudice-regulation. That is, if a reduction in defensive responding increases the success of anti-bias interventions and prejudice-regulation, what might matter most for the long-term persistence of bias reduction is the success of these strategies, not necessarily long-term changes in beliefs about bias and discrimination. Nevertheless, future research should explore the long-term effects of motivated reasoning processes and reductions in

defensive responding in this context, and its relationship to the success of anti-bias interventions and strategies for prejudice-regulation.

Finally, there may be alternative explanations for the observed pattern of effects other than the proposed mechanism identified here (i.e., threat to egalitarians self-image). For example, it is possible that the motivated rejection of implicit racial bias and feedback has less to do with the threat to one's self image or the lack of efficacy to control the influence of unconscious bias, and more to do with aversion to the public policy implications of subtle forms of racial bias (Campbell & Kay, 2014). This may be particularly true for political conservatives who are traditionally skeptical of public policies designed to reduce racial discrimination (Sniderman, Piazza, Tetlock, & Kendrick, 1991). Additional work suggests that the threat to an egalitarian self-image may not be limited to one's self-concept, but may extend to collective identities (e.g., racial group membership; Adams, Tormala, & O'Brien, 2006). Evidence that members of one's social group are complicit in or directly responsible for discrimination against target groups may elicit feelings of collective guilt (e.g., Doosje, Branscombe, Spears, & Manstead, 1998), or otherwise threaten the esteem of one's social group membership and the legitimacy of existing hierarchy (O'Brien, 2002). Alternatively, a lack of efficacy to control the automatic expression of prejudice may also promote denial and behavioral avoidance, as has been observed in other contexts (e.g., health domain; Witte & Allen, 2000) and in the present research. In either case, additional work should evaluate these alternative mechanisms to better understand the relative contributions of each in producing the observed pattern of effects.

Importantly, the psychological mechanisms responsible for the motivated rejection of implicit bias evidence and feedback may be multiple and diverse. Strategies to attenuate motivated reasoning processes and improve awareness should directly address these underlying factors. In this work, I focused primarily on reducing the threat to one's self-image and a perceived lack of control in the expression of unconscious bias. However, alternative interventions may also be effective. For example, adopting an alternative frame for describing public policies that are intended to help racial minorities (e.g., emphasizing equality of opportunity) may attenuate aversion to the policy remedies implied by the existence of implicit racial bias. Similarly, providing people with the opportunity to affirm the legitimacy of relevant social systems or the positivity of one's social group may be a more constructive approach then the self-affirmation intervention used here.

Conclusion

In summary, the current work illuminates the psychological factors associated with the motivated rejection of feedback and evidence of implicit racial bias, and its consequences for endorsement of racial stereotypes, attitudes about public policy, and the motivation to control prejudice responding and participate in anti-prejudice interventions. I also tested two strategies to attenuate these motivated reasoning processes, and find strong evidence for the utility of pairing personalized feedback with a "collective bias intervention" in order to reduce defensive responding. Indeed, the threat of being labeled a racist, and the discomfort associated with the prospect of harboring unconscious forms of racial bias is inconsistent with White Americans' egalitarian intentions and self-image, as well as the perceived controllability of prejudice responding. As such, the motivated

rejection of implicit bias may constitute a major obstacle to the type of self-awareness necessary for the effective regulation of prejudice responding. Understanding these obstacles to self-awareness and identifying strategies to overcome them is necessary to remedy persistent racial inequalities that have no place in an egalitarian society.

References

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

Abramowitz, A. I., & Saunders, K. L. (1998). Ideological realignment in the US electorate. *Journal of Politics*, *60*, 634-652.

Adams, G., Tormala, T. T., & O'Brien, L. T. (2006). The effect of self-affirmation on perception of racism. *Journal of Experimental Social Psychology*, *42*(5), 616-626.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market.*Handbook of labor economics*, *3*, 3143-3259.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of personality and social psychology*, *91*(4), 652.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of personality and social psychology*, *84*(4), 738.

Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., & Amadori, A. (2008). Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, *29*(3), 369-387.

Ayres I. 2001. Pervasive Prejudice: Unconventional Evidence of Race and Gender Discrimination. Chicago: Univ. Chicago Press

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*(1), 4.

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. Delacorte Press.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of personality and Social Psychology*, *65*(2), 272.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für experimentelle Psychologie*, *48*(2), 145-160.

Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood.*Psychological science*, *17*(1), 53-58.

Banaji, M. R. (2001). Ordinary prejudice. *Psychological Science Agenda, American Psychological Association*, *14*(Jan-Feb).

Bartlett, K. T. (2009). Making good on good intentions: The critical role of motivation in reducing implicit workplace discrimination. *Virginia Law Review*, 1893-1972.

Baumeister, R. F., & Tice, D. M. (1985). Self-esteem and responses to success and failure: Subsequent performance and intrinsic motivation.*Journal of Personality*, *53*(3), 450-467.

Bendick, M., & Nunes, A. P. (2012). Developing the research basis for controlling bias in hiring. *Journal of Social Issues*, *68*(2), 238-262.

Bendick, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring employment discrimination through controlled experiments. *The Review of Black Political Economy*, *23*(1), 25-48.

Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of personality and social psychology*, *99*(2), 248.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, *20*(3), 351-368.

Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of personality and social psychology*, *99*(2), 248.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, *94*(4), 991-1013.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 359-374).

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice.*Personality and Social Psychology Review*, *6*(3), 242-261.

Blank, R. M. (2001). An overview of trends in social and economic well-being, by race. *America becoming: Racial trends and their consequences*, *1*, 21-39.

Bobo, L., & Kluegel, J. R. (1993). Opposition to race-targeting: Self-interest, stratification ideology, or racial attitudes?. *American Sociological Review*, 443-464.

Babe, L. D. (2001). Racial attitudes and relations at the close of the twentieth century. Smelser, N.J., Wilson, W.J, Mithcell, F (eds), America Becoming: Racial Trends and Their Consequence (VOL 1.). DC: National Academy Press.

Bobo, L., Kluegel, J. R., & Smith, R. A. (1997). Laissez-faire racism: The crystallization of a kinder, gentler, antiblack ideology. *Racial attitudes in the 1990s: Continuity and change*, *15*, 23-5.

Borgida, E., & Fiske, S. T. (Eds.). (2008). *Beyond common sense: Psychological science in the courtroom*. John Wiley & Sons.

Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., & Vaslow, J. B. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational behavior and human decision processes*, *81*(1), 72-97.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, *6*(1), 3-5.

Burke, P. J., & Harrod, M. M. (2005). Too much of a good thing?. *Social Psychology Quarterly*, *68*(4), 359-374.

Burke, P.J. (1991). Identity processes and social stress. *American Sociological Review, 56*, 836-849.

Burke, P.J (2006). Identity change. *Social Psychology Quarterly*, 69, 81-96.

Cast, A. D., & Burke, P. J. (2002). A theory of self-esteem. *Social forces*,*80*(3), 1041-1068.

Collange, J., Fiske, S. T., & Sanitioso, R. (2009). Maintaining a positive self-image by stereotyping others: Self-threat and the stereotype content model.*Social cognition*, *27*(1).

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals.*Journal of personality and social psychology*, *83*(6), 1314.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization.*Journal of personality and social psychology*, *82*(3), 359.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review.*Psychological Bulletin*, *87*(3), 546.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*(3), 316-328.

De Houwer, J. (2006). What are implicit measures and why are we using them. *The handbook of implicit cognition and addiction*, 11-28.

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. *Affect, cognition, and stereotyping: Interactive processes in group perception*, 317-344.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of personality and social psychology*,*82*(5), 835.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267-1278.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, *56*(1), 5.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*,*60*(6), 817.

Dixon, J., Durrheim, K., & Tredoux, C. (2007). Intergroup contact and attitudes toward the principle and practice of racial equality. *Psychological Science*, *18*(10), 867-872.

Doleac, J. L., & Stein, L. C. (2013). The visible hand: Race and online market outcomes. *The Economic Journal*, *123*(572), F469-F492.

Dovidio, J. F., & Gaertner, S. L. (1986). Prejudice, discrimination, and racism: Historical trends and contemporary approaches. In J. F. Dovidio & S. L.

Gaertner (Eds.), Prejudice, discrimination, and racism (pp. 1-34). New York: Academic Press.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. Eberhardt & S. T. Fiske (Eds.), Confronting racism: The problem and the response (pp. 1-32). Newbury Park, CA: Sage.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. Psychological Science, 11, 319-323.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 36, pp. 1-51). San Diego, CA: Academic Press.

Duguid, M. M., & Thomas-Hunt, M. C. (2015). Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes.*Journal of Applied Psychology*, *100*(2), 343.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316-326.

Eberhardt, J. L., & Fiske, S. T. (Eds.). (1998). *Confronting racism: The problem and the response*. Sage Publications.

Eibach, R. P., & Keegan, T. (2006). Free at last? Social dominance, loss aversion, and White and Black Americans' differing assessments of racial progress. *Journal of personality and social psychology*, *90*(3), 453.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline?. *Journal of personality and social psychology*,*69*(6), 1013.

Federico, C. M., & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: the antecedents and consequences of" principled objections" to affirmative action. *Journal of personality and social psychology*, *82*(4), 488.

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

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*(2), 117-140.

Fisher, E. L., & Borgida, E. (2012). Intergroup disparities and implicit bias: A commentary. *Journal of Social Issues*, *68*(2), 385-398.

Frantz, C. M., Cuddy, A. J., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*(12), 1611-1624.

Gaertner, S. L., & Dovidio, J. F. (1986). Prejudice, discrimination, and racism: Problems, progress, and promise. In J. F. Dovidio & S. L. Gaertner (Eds.), Prejudice, discrimination, and racism (pp. 315-332). New York: Academic Press.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious?. *Consciousness and cognition*, *15*(3), 485-499.

Gilens, M. (1995). Racial attitudes and opposition to welfare. *The Journal of Politics*, *57*(04), 994-1014.

Gilens, M. (1999). Why Americans hate welfare.

Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice.*Journal of Experimental Social Psychology*, *44*(1), 164-172.

Gomez, B. T., & Wilson, J. M. (2006). Cognitive heterogeneity and economic voting: A comparative analysis of four democratic electorates. *American Journal of Political Science*, *50*(1), 127-145.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine*, *22*(9), 1231-1238.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*(1), 4.

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations.*California Law Review*, *94*(4), 945-967.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, *97*(1), 17.

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy*, *9*(1), 241-253.

Hansen F. 2003. Diversity's business case: doesn't add up. Workforce 824:28–32

Hawkins, C. B., & Nosek, B. A. (2012). Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin*, *38*(11), 1437-1452.

Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, *23*(2), 253-283.

Hillard, A. L., Ryan, C. S., & Gervais, S. J. (2013). Reactions to the implicit association test as an educational tool: A mixed methods study. *Social Psychology of Education*, *16*(3), 495-516.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*,*31*(10), 1369-1385.

Howell, J.L., Collisson, B.D., Crysel, L., Garrido, C.O., Newell, S.M. Cottrell, C.A., Smith, C.T.S. & Shepperd, J.A. (2013). Managing the threat of impending implicit attitude feedback. Social Psychological and Personality Science, 4, 714-720

Howell, J.L., Gaither, S.E., & Ratliff, K.A. (2015). Caught in the Middle: Defensive Responses to IAT feedback among Whites, Blacks, and Biracial Black/Whites. Social Psychological and Personality Science.

Howell, J.L., & Shepperd, J.A. (2012). Reducing information avoidance through affirmation. Psychological Science, 23, 141-145.

Huddy, L., & Feldman, S. (2006). Worlds apart: Blacks and whites react to Hurricane Katrina. *Du Bois Review*, *3*(01), 97-113.

Huddy, L., & Feldman, S. (2009). On assessing the political effects of racial prejudice. *Annual Review of Political Science*, *12*, 423-447.

Hurwitz, J., & Peffley, M. (1997). Public perceptions of race and crime: The role of racial stereotypes. *American journal of political science*, 375-401.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies.*American sociological review*, *71*(4), 589-617.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in organizational behavior*, *29*, 39-69.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political psychology*, *25*(6), 881-919.

Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, *27*(2), 254-263.

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of" affirmative action". *California Law Review*, *94*(4), 1063-1118.

Katz, I., Wackenhut, J. & Hass, G.R. (1986)Racial ambivalence, value duality, and behavior. Dovidio, J.F., Gaertner, S.L. (EdS), Prejudice, discrimination, and racism. , (pp. 35-59). San Diego, CA, US: Academic Press, xiii, 337 pp

Kay, A. C., Gaucher, D., Napier, J. L., Callan, M. J., & Laurin, K. (2008). God and the government: testing a compensatory control mechanism for the support of external systems. *Journal of personality and social psychology*,*95*(1), 18.

Kinder, D. R., & Mendelberg, T. (2000). Individualism reconsidered: Principles and prejudice in contemporary American opinion. *Racialized politics: The debate about racism in America*, 44-74.

Kinder, D. R., & Sanders, L. M. (1996). Divided by color.

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of personality and social psychology*, *40*(3), 414.

Kluegel, J. R. (1990). Trends in whites' explanations of the black-white gap in socioeconomic status, 1977-1989. *American Sociological Review*, 512-525.

Krieger, L. H. (2008). Behavioral realism in law: reframing the discussion about social science's place in antidiscrimination law and policy. *See Borgida & Fiske*, *2008*, 383-98.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*,*108*(3), 480.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., ... & Frazier, R. S. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765.

Lauritsen, J. L., & Sampson, R. J. (1998). Minorities, crime, and criminal justice. *The handbook of crime and punishment*, *58*, 65-70.

Legault, L., Green-Demers, I., Grant, P., & Chung, J. (2007). On the Self-Regulation of Implicit and Explicit Prejudice A Self-Determination Theory Perspective. *Personality and Social Psychology Bulletin*, *33*(5), 732-749.

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages how motivational interventions can reduce (but also increase) prejudice. *Psychological Science*, 0956797611427918.

Legault, L., Green-Demers, I., & Eadie, A. L. (2009). When internalization leads to automatization: The role of self-determination in automatic stereotype suppression and implicit prejudice regulation. *Motivation and Emotion*, *33*(1), 10-24.

Leslie, L. M., King, E. B., Bradley, J. C., & Hebl, M. R. (2008). Triangulation across methodologies: All signs point to persistent stereotyping and discrimination in organizations. *Industrial and Organizational Psychology*,*1*(04), 399-404.

Levinson, J. D., Cai, H., & Young, D. (2009). Guilty by implicit racial bias: The guilty/not guilty Implicit Association Test. *Ohio State Journal of Criminal Law, Forthcoming*.

MacDonald, J. M. (2001). Analytic methods for examining race and ethnic disparity in the juvenile courts. *Journal of Criminal Justice*, *29*(6), 507-519.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, *44*(1), 1-23.

Mauer, M. (1999). The crisis of the young African American male and the criminal justice system. *Impacts of incarceration on the African American family*, 199.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale.

McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*(4), 289-354.

Mendelberg, T. (2001). *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press.

Monahan, J., Walker, L., & Mitchell, G. (2008). Contextual evidence of gender discrimination: The ascendance of" social frameworks". *Virginia Law Review*, 1715-1749.

Monteith, M. J., Deneen, N. E., & Tooman, G. D. (1996). The effect of social norm activation on the expression of opinions concerning gay men and Blacks. *Basic and Applied Social Psychology*, *18*(3), 267-288.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. Journal of Personality and Social Psychology, 64, 198-210.

Monteith, M. J., & Mark, A. Y. (2009). The self-regulation of prejudice.*Handbook of prejudice, stereotyping, and discrimination*, 507-523.

Monteith, M. J., Spicer, C. V., & Tooman, G. D. (1998). Consequences of stereotype suppression: Stereotypes on AND not on the rebound. *Journal of Experimental Social Psychology*, *34*(4), 355-377.

Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology*, *75*(4), 901.

Monteith, M. J., & Walters, G. L. (1998). Egalitarianism, moral obligation, and prejudice-related personal standards. *Personality and Social Psychology Bulletin*, *24*(2), 186-199.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice reduction efforts. Journal of Personality and Social Psychology, 65, 469-485. Monteith, M. J. (1996). Affective reactions to prejudice-related discrepant responses: The impact of standard salience. Personality and Social Psychology Bulletin, 22, 48-59.

Morsella, E., & Bargh, J. A. (2011). Unconscious action tendencies: Sources of 'un-integrated'action. *The handbook of social neuroscience*, 335-347.

National Research Council (2004). Blank, R. M., Dabady, M., & Citro, C. F. (Eds.). *Measuring racial discrimination*. National Academies Press.

Norton, M. I., Sommers, S. R., Apfelbaum, E. P., Pura, N., & Ariely, D. (2006). Color blindness and interracial interaction playing the political correctness game. *Psychological Science*, *17*(11), 949-953.

Norton, M. I., Vandello, J. A., Biga, A., & Darley, J. M. (2008). Colorblindness and diversity: Conflicting goals in decisions influenced by race. *Social Cognition*, *26*(1), 102.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the implicit association test. *Experimental psychology*, *54*(1), 14-29.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*(4), 565.

O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm No Bigot: How Prejudiced White Americans Maintain Unprejudiced Self-Images. *Journal of Applied Social Psychology*,*40*(4), 917-946.

Oliver, M. L., & Shapiro, T. M. (1995). Black wealth/white wealth.

Oppler, S. H., Campbell, J. P., Pulakos, E. D., & Borman, W. C. (1992). Three approaches to the investigation of subgroup bias in performance measurement: Review, results, and conclusions. *Journal of Applied Psychology*, *77*(2), 201.

Page Benjamin, I., & Shapiro Robert, Y. (1992). The rational public: Fifty years of trends in Americans' policy preferences.

Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The Annals of the American Academy of Political and Social Science*, *609*(1), 104-133.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual review of psychology*, *60*, 339-367.

Paolacci, G., & Chandler, J. (2014). Inside the turk understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184-188.

Peffley, M., & Hurwitz, J. (2010). *Justice in America: The separate realities of Blacks and Whites*. Cambridge University Press.

Penner, L. A., Eggly, S., Griggs, J. J., Underwood, W., Orom, H., & Albrecht, T. L. (2012). Life-Threatening Disparities: The Treatment of Black and White Cancer Patients. *Journal of Social Issues*, *68*(2), 328-357.

Pettigrew, T. F. (2003). Peoples under threat: Americans, Arabs, and Israelis.*Peace and Conflict: Journal of Peace Psychology*, *9*(1), 69.

Plant, E. A., & Butz, D. A. (2006). The causes and consequences of an avoidance-focus for interracial interactions. *Personality and Social Psychology Bulletin*, *32*(6), 833-846.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of personality and social psychology*,*75*(3), 811.

Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash?. *Journal of Experimental Social Psychology*, *37*(6), 486-501.

Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *16*(3), 180-183.

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes & Intergroup Relations*, *6*(2), 187-200.

Plant, E. A., Devine, P. G., & Peruche, M. B. (2010). Routes to positive interracial interactions: Approaching egalitarianism or avoiding prejudice.*Personality and Social Psychology Bulletin*, *36*(9), 1135-1147.

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European review of social psychology*, *17*(1), 271-320.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, *67*(4), 741.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers*, *36*(4), 717-731.

Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology*, 299-328.

Reyna, C., Tucker, A., Korfmacher, W., & Henry, P. J. (2005). Searching for common ground between supporters and opponents of affirmative action.*Political Psychology*, *26*(5), 667-682.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*(3), 417-423.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay effects of interracial contact on executive function. *Psychological Science*, *14*(3), 287-290.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: a new meta-analysis. *Journal of Applied Psychology*, *88*(4), 694.

Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes & Intergroup Relations*, *5*(2), 133-150.

Rudman, L. A., Dohn, M. C., & Fairchild, K. (2007). Implicit self-esteem compensation: automatic threat defense. *Journal of personality and social psychology*, *93*(5), 798.

Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not Black and White: Racial bias in the decision to shoot in a multiethnic context.*Journal of Social Issues*, *68*(2), 286-313.

Saucier, D. A. (2002). Self-reports of racist attitudes for oneself and for others. *Psychologica belgica*, *42*(1-2), 99-105.

Scheck, J. (2004, October 28). Expert witness: Bill Bielby helped launch an industry–suing employers for unconscious bias. Retrieved from http:// www.law.com/jsp/PubArticle.jsp?id900005417471

Schuman, H., Steeh, C., & Bobo, L. (1985). Racial trends in America: Trends and interpretations.

Schwartz, S. H. (1977). Normative influences on altruism. *Advances in experimental social psychology*, *10*, 221-279.

Sears, D. O., & Henry, P. J. (2003). The origins of symbolic racism. *Journal of personality and social psychology*, *85*(2), 259.

Sears, D. O., Van Laar, C., Carrillo, M., & Kosterman, R. (1997). Is it really racism?: The origins of white Americans' opposition to race-targeted policies.*The Public Opinion Quarterly*, *61*(1), 16-53.

Sears, D. O., Henry, P. J., & Kosterman, R. (2000). Egalitarian values and contemporary racial politics (pp. 75–117).

Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. Group Processes and Intergroup Relations, 6, 171–186.

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, *2*(2), 895-908.

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self–Affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119-123.

Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, *26*(9), 1046-1058.

Sidanius, J. P., & Pratto, F. F.(1999). Social dominance: An intergroup theory of social hierarchy and oppression.

Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance?. *Journal of personality and social psychology*, *70*(3), 476.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003). Committee on understanding and eliminating racial and ethnic disparities in health care. Unequal treatment:

confronting racial and ethnic disparities in health care. In*National Academy of Science* (Vol. 180, p. 191).

Sniderman, P. M., & Piazza, T. (1993). The scar of race. *Cambridge: Belknap/Harvard Univ. Press.*

Sniderman, P. M., Carmines, E. G., Layman, G. C., & Carter, M. (1996). Beyond race: Social justice as a race neutral ideal. *American Journal of Political Science*, 33-55.

Sommers, S. R., & Norton, M. I. (2006). Lay theories about White racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations*, *9*(1), 117-138.

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

Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Duinn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*(11), 1139-1152.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology*, *21*, 261-302.

Strauman, T. J., & Higgins, E. T. (1987). Automatic activation of self-discrepancies and emotional syndromes: when cognitive structures influence affect. *Journal of personality and social psychology*, *53*(6), 1004.

Stryker, S., & Burke, P. J. (2000). The past, present, and future of an identity theory. *Social psychology quarterly*, 284-297.

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

Swann Jr, W. B., & Bosson, J. K. (2008). Identity negotiation. *Handbook of Personality: Theory and Research*, 448.

Swann, W. B. (1987). Identity negotiation: where two roads meet. *Journal of personality and social psychology*, *53*(6), 1038.

Swann Jr, W. B. (2005). The self and identity negotiation. *Interaction Studies*,*6*(1), 69-83.

Tesler, M., & Sears, D. O. (2010). *Obama's race: The 2008 election and the dream of a post-racial America*. University of Chicago Press.

Tesler, M. (2012). The spillover of racialization into health care: How President Obama polarized public opinion by racial attitudes and race.*American Journal of Political Science*, *56*(3), 690-704.

Tesler, M. (2013). The return of old-fashioned racism to White Americans' partisan preferences in the early Obama era. *The Journal of Politics*, *75*(01), 110-123.

Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. 939 (U.S. 2015).

Tormala, Z. L., Clarkson, J. J., & Petty, R. E. (2006). Resisting persuasion by the skin of one's teeth: the hidden success of resisted persuasive messages. *Journal of personality and social psychology*, *91*(3), 423.

Vera, H., Feagin, J. R., & Gordon, A. (1995). Superior intellect?: Sincere fictions of the white self. *Journal of Negro Education*, 295-306.

Virtanen, S. V., & Huddy, L. (1998). Old-fashioned racism and new forms of racial prejudice. *The journal* Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction.*Journal of personality and social psychology*, *78*(4), 690.

Vorauer, J. D., Main, K. J., & O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of personality and social psychology*, *75*(4), 917.

Walker, S., Spohn, C., & DeLone, M. (2004). The Color of Justice: Race, Ethnicity, and Crime in America; Thomson/Wadsworth. *Belmont, CA*.

Western, B. (2006). *Punishment and inequality in America*. Russell Sage Foundation.

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health education & behavior*, *27*(5), 591-615.

Yates, J., & Fording, R. (2005). Politics and state punitiveness in black and white. *Journal of Politics*, *67*(4), 1099-1121.

Yinger, J. (1993). Access denied, access constrained: results and implications of the 1989 housing discrimination study. *Clear and convincing evidence: Measurement of discrimination in America*, 69-112.

Yinger, J. (1998). Evidence on discrimination in consumer markets. *The Journal of Economic Perspectives*, *12*(2), 23-40.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of consumer research*,*37*(2), 197-206.

Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: the role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, *90*(3), 553.

Zuwerink, J. Z., Devine, P. G., Monteith, M. J., & Cook, D. A. (1996). Prejudice toward blacks: With and without compunction?. *Basic and Applied Social Psychology*, *18*(2), 131-150.

Tables

Table 1.1

Means, SD, and alpha for measures used in Pilot 1b

Variable (Range)	Mean	SD	Alpha
Reading Comprehension (0-4)	3.76	.54	--
Perceived Believability (0-1)	.74	.19	--
Perceived Accuracy (0-1)	.40	.09	--
Perceived Deception (0-1)	.26	.44	--
Shortened Motivated Reasoning Battery (0-1)	.54	.14	.82
Self-Perceived Implicit Racial Bias (0-1)	.49	.25	.85
Perceptions of Social Science (0-1)	.59	.21	.86
Belief in Racial Prejudice and Discrimination (0-1)	.60	.20	.90

Table 1.2

Correlations between all continuous variables used in Study 1

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. T1 Racial Resentment	.56	.26	--											
2. T1 Attitudes Toward Blacks	.29	.16	.56**	--										
3 T1 Internal-Ext Egal. Motivation	.67	.20	-.35**	-.66**	--									
4. T1 SDO	.22	.20	.48**	.80**	-.59**	--								
5. T1 System Justification	.43	.19	.33**	.35**	-.29**	.37**	--							
6. T2 Self-Perceived Implicit Bias	.56	.23	-.01	.14**	-.28**	.12*	.08	--						
7. T2 Perceptions of Social Science	.61	.18	-.38**	-.39**	-.28**	-.45**	-.07	-.05	--					
8. T2 Belief in Prejudice & Discrim.	.67	.17	-.49**	-.43**	.32**	-.43**	-.3**	.38**	.31**	--				
9. T2 General Egal. Motivation	.52	.14	-.24**	-.22**	.15**	-.17**	.03	.00	.19**	.23**	--			
10. T2 Anti-Bias Intervention Attitudes	.64	.26	-.39**	-.45**	.42**	-.45**	-.17**	.13*	.30**	.62**	.29**	--		
11. T2 Racial Stereotyping	.42	.26	.71**	.68**	-.45**	.66**	.38**	.04	-.43**	-.56**	-.12*	-.49**	--	
12. T2 Public Policy Attitudes	.57	.19	-.69**	-.60**	.45**	-.57**	-.32**	.03	.39**	.60**	.21**	.51**	-.76**	--

Note. †$p < .10$. *$p < .05$. **$p < .01$.

Table 2.1
Means, SD, and alpha for measures used in Pilot 2

Variable	Mean	SD	Alpha
Perceived Believability	.66	.72	--
Perceived Accuracy	.61	.51	--
Perceived Deception	.40	.49	--
Overall Negative Affect	.22	.22	.97
Overall Positive Affect	.53	.24	.92
Depressed	.24	.29	.92
Discomfort	.22	.23	.94
Other-Directed Negative	.20	.22	.88
Self-Directed Negative	.23	.26	.95
Full Motivated Reasoning Battery	.53	.15	.87
Self-Perceived Implicit Racial Bias	.43	.25	.86
Perceptions of Social Science	.59	.20	.87
Perceptions of IAT	.31	.20	.86
Belief in Racial Prejudice and Discrimination	.60	.20	.88

Table 2.2

Correlations between all continuous variables used in Pilot 2

Variables	1	2	3	4	5	6
1. T2 Overall Negative	--					
2. T2 Overall Positive	-.38**	--				
3. T2 Self-Perceived Implicit Bias	.29**	-.05	--			
4. T2 Perceptions of Social Science	.02	.04	.06	--		
5. T2 Perceptions of IAT	.12*	.11†	.39**	.17**	--	
6. T2 Belief in Prejudice & Discrim.	.13**	-.10	.40**	.30**	.27**	--

Note. †$p < .10$. *$p < .05$. **$p < .01$.

Table 2.3
Inter-rater agreement for open-ended content coding of self-reported belief in the use of deception, prior to consensus, for Pilot 2

What is the purpose of this Study? (Category)	Inter-rater Agreement (Kappa)
(1) Response is a Reaction to IAT	.39
(2) Purpose is to Study Race and Inform	.37
(3) Purpose is to Manipulate Beliefs on Race	.57
(4) Unspecified Use of Deception	.40

Table 2.4
Correlations between all continuous variables used in Study 2

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. T1 Racial Resentment	.54	.27	--															
2. T1 Attitudes Toward Blacks	.70	.17	-.003	--														
3. T1 Internal-Ext Egal. Motivation	.65	.18	.30**	-.05	--													
4. T1 SDO	.26	.19	-.41**	.03	-.49**	--												
5. T1 System Justification	.44	.20	-.32**	-.001	-.17**	.26**	--											
6. T2 Overall Negative	.26	.24	-11†	.03	-.05	-.06	-.10*	--										
7. T2 Overall Positive	.44	.28	-.08	.02	.02	.14*	.13*	-.40**	--									
8. T2 Self-Perceived Implicit Bias	.32	.17	-.03	.03	-.15**	-.01	.06	.18**	-.07	--								
9. T2 Perceptions of Social Science	.65	.18	-28**	-.20**	.21**	-.39**	-.05	.04	-.08	.12*	--							
10. T2 Perception of IAT	.35	.23	-03	-.11*	-.02	-.08	.02	-.00	.20**	.37**	.26**	--						
11. T2 Belief in Prejudice & Discrim.	.52	.17	-.50**	.04	.27**	-.48**	-.32**	.05	.01	.34**	.29**	.37**	--					
12. T2 Race IAT	.61	.09	-.08	.06	.004	-.01	.02	.08	-.08	.09	-.001	-.05	-.03	--				
13. T2 General Egal. Motivation	.51	.15	.36**	-.02	.13*	-.31**	-.07	.07	-.09†	.06	.18**	.12*	.30**	-.03	--			
14. T2 Anti-Bias Intervention Attitudes	.57	.26	.43**	-.03	.40**	-.47**	-.24**	.05	.11†	.15*	.32**	.34**	.66**	-.05	.40**	--		
15. T2 Racial Stereotyping	.43	.26	-.73**	.02	-.36**	.57**	.32**	-.08	.13*	.03	-.44**	-.09	-.57**	.08	-.33**	-.49**	--	
16. T2 Public Policy Attitudes	.49	.18	.62**	.01	.39**	-.61**	-.34**	.06	-.08	.05	.44**	.13*	.55**	-.09	.33**	.54**	-.77**	--

Note. †$p < .10$. *$p < .05$. **$p < .01$.

Table 3.1
Means, SD, and alpha for measures used in Pilot 3

Variable	Mean	SD	Alpha
Perceived Believability	.54	.20	--
Perceived Accuracy	.70	.21	--
Perceived Deception	.46	.50	--
Overall Negative Affect	.32	.25	.97
Overall Positive Affect	.38	.26	.93
Depressed	.31	.31	.92
Discomfort	.31	.26	.93
Other-Directed Negative	.23	.23	.86
Self-Directed Negative	.37	.31	.96
Full Motivated Reasoning Battery	.47	.12	.81
Self-Perceived Implicit Racial Bias	.40	.21	.86
Perceptions of Social Science	.41	.19	.83
Perceptions of IAT	.36	.26	.92
Belief in Racial Prejudice and Discrimination	.63	.20	.89

Table 3.2

Correlations between all continuous variables used in Pilot 3

Variables	1	2	3	4	5	6
1. T2 Overall Negative	--					
2. T2 Overall Positive	-.35**	--				
3. T2 Self-Perceived Implicit Bias	.39**	-.10*	--			
4. T2 Perceptions of Social Science	-.05	.22**	-.09†	--		
5. T2 Perceptions of IAT	.27**	-.02	.62**	-.13*	--	
6. T2 Belief in Prejudice & Discrim.	.36**	-.18**	.49**	-.25**	.41**	--

Note. †$p < .10$. *$p < .05$. **$p < .01$.

Table 3.3
Means and SD for measures used in Study 3a

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. T1 Racial Resentment	.55	.26	--										
2. T1 Attitudes Toward Blacks	.29	.18	.58**	--									
3. T1 Internal-Ext Egal. Motivation	.69	.19	-.26**	-.62**	--								
4. T1 SDO	.24	.16	.39**	.77**	-.54**	--							
5. T1 System Justification	.41	.19	.33**	.31**	-.21**	.34**	--						
6. T2 Overall Negative	.30	.24	-.18**	-.11*	.001	-.05	-.12*	--					
7. T2 Overall Positive	.41	.28	.25**	.23**	-.07	.22**	.28**	-.34**	--				
8. T2 Self-Perceived Implicit Bias	.34	.18	-.09**	.15**	-.29**	.18**	.02	.24**	-.07	--			
9. T2 Anti-Bias Intervention Attitudes	.60	.25	-.33**	-.45**	.29**	-.34**	-.15**	.26**	-.02	.23**	--		
10. T2 Racial Stereotyping	.37	.24	.65**	.75**	-.40**	.60**	.35**	-.16**	.33**	.07	-.43**	--	
11. T2 Public Policy Attitudes	.58	.19	-.66**	-.65**	.37**	-.54**	-.35**	.15**	-.25**	.05	.45**	-.69	--

Note. †$p < .10$. *$p < .05$. **$p < .01$

Table 3.4
Means and SD for measures used in Study 3b

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. T1 Racial Resentment	.53	.27	--									
2. T1 Internal-Ext Egal. Motivation	.69	.18	-.35**	--								
3. T1 SDO	.22	.16	.45**	-.51**	--							
4. T1 System Justification	.45	.64	.41**	-.20**	.39**	--						
5. T2 Overall Negative	.29	.24	-.17**	-.06	-.09†	-.24**	--					
6. T2 Overall Positive	.45	.29	.23**	-.12*	.16**	.25**	-.30**	--				
7. T2 Self-Perceived Implicit Bias	.33	.18	.06	-.35**	.12*	-.02	.30**	-.03	--			
8. T2 Perceptions of Social Science	.49	.16	-,29**	.20**	-.34**	-.06	-.02	-.05	.06	--		
9. T2 Perception of IAT	.35	.23	.08	-.28**	.08	.04	.07	.22**	.45**	.19**	--	
10. T2 Belief in Prejudice & Discrim.	.58	.16	-.49**	.10*	-.33**	-.42**	.30**	-.16**	.40**	.25**	.31**	--

Note. †$p < .10$. *$p < .05$. **$p < .01$

Figures

Figure 1.1
Main effect of experimental condition on motivated reasoning for Pilot 1a

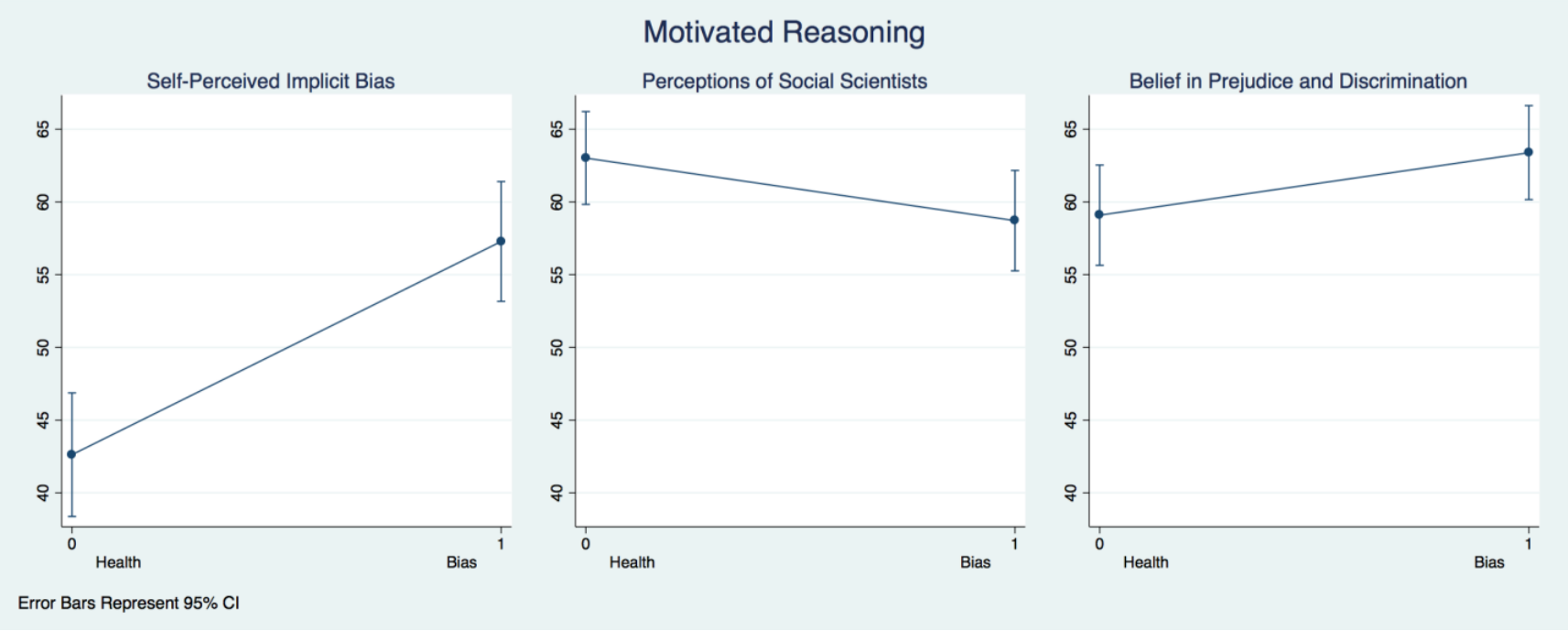


Figure 1.2
Main effect of experimental condition on motivated reasoning for Pilot 1b

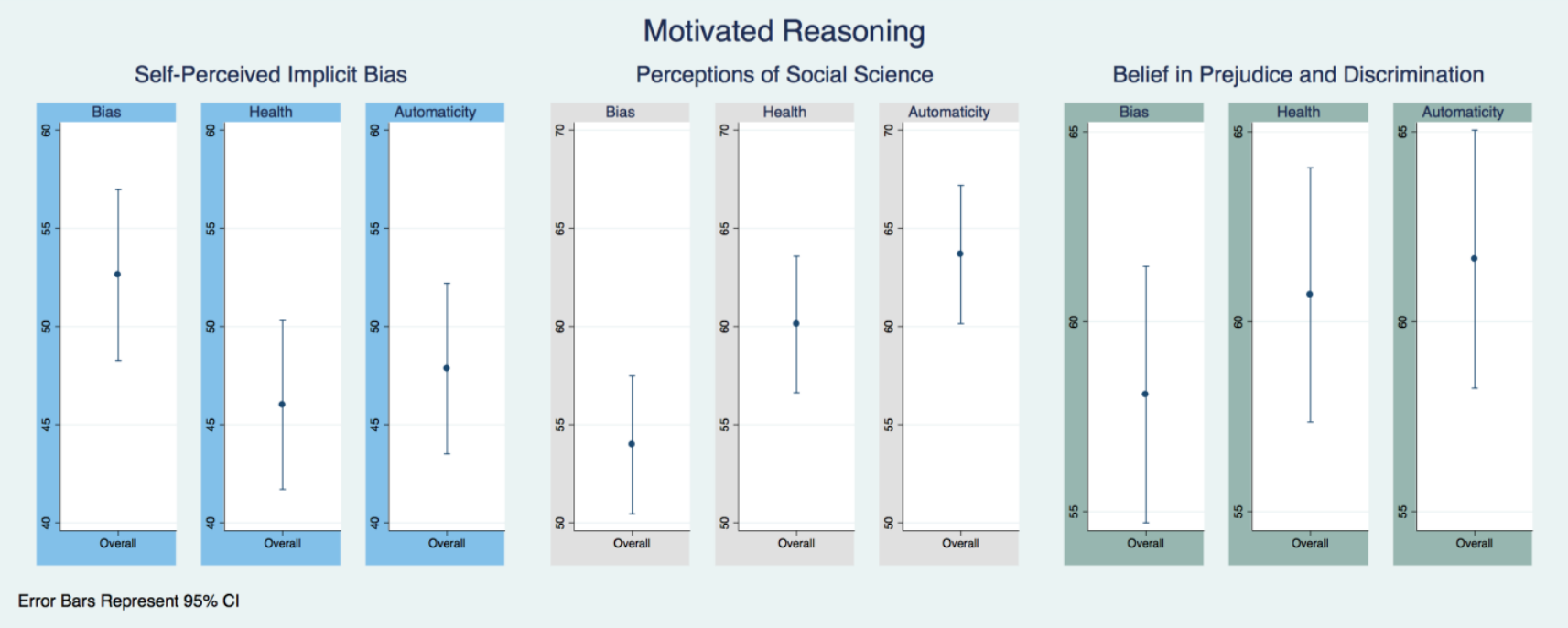


Figure 1.3
Coefficients plot for effect of T1 independent variables on motivated reasoning in Study 1

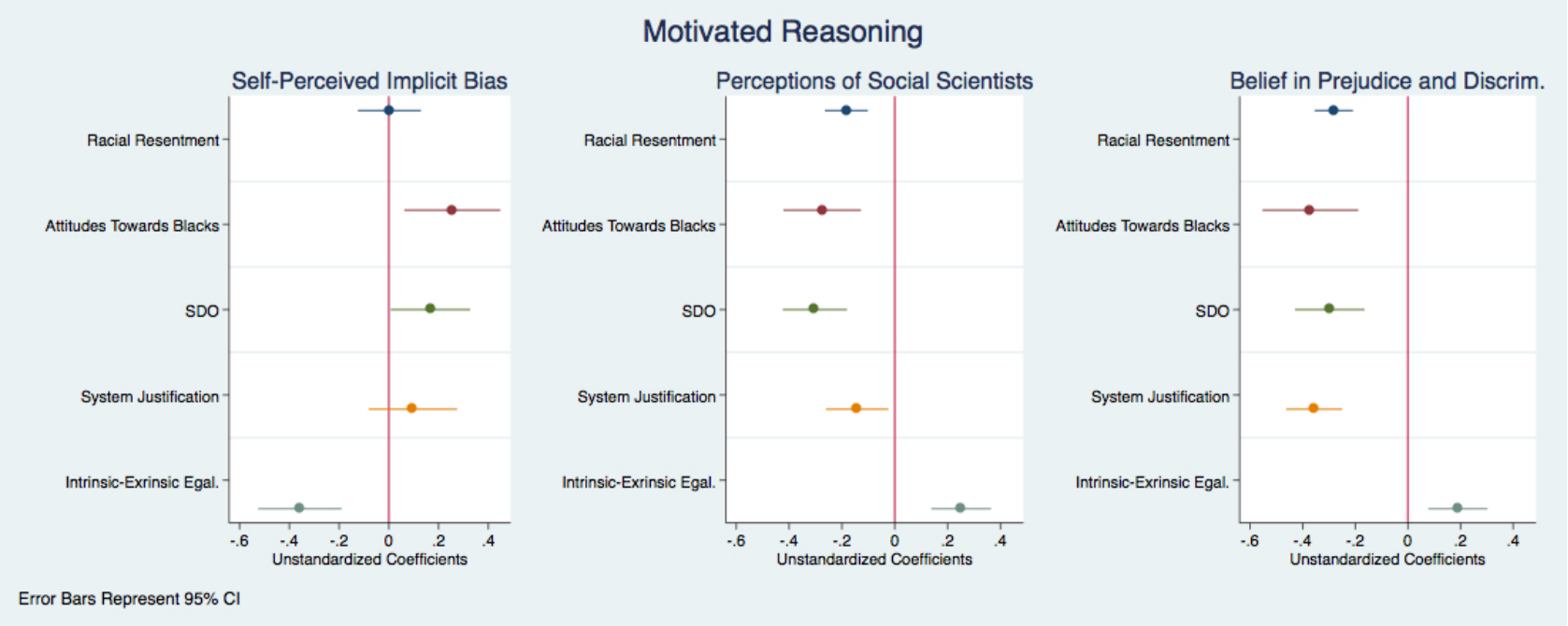


Figure 1.4

Coefficients plot for effect of T1 independent variables on prejudice-regulation in Study 1

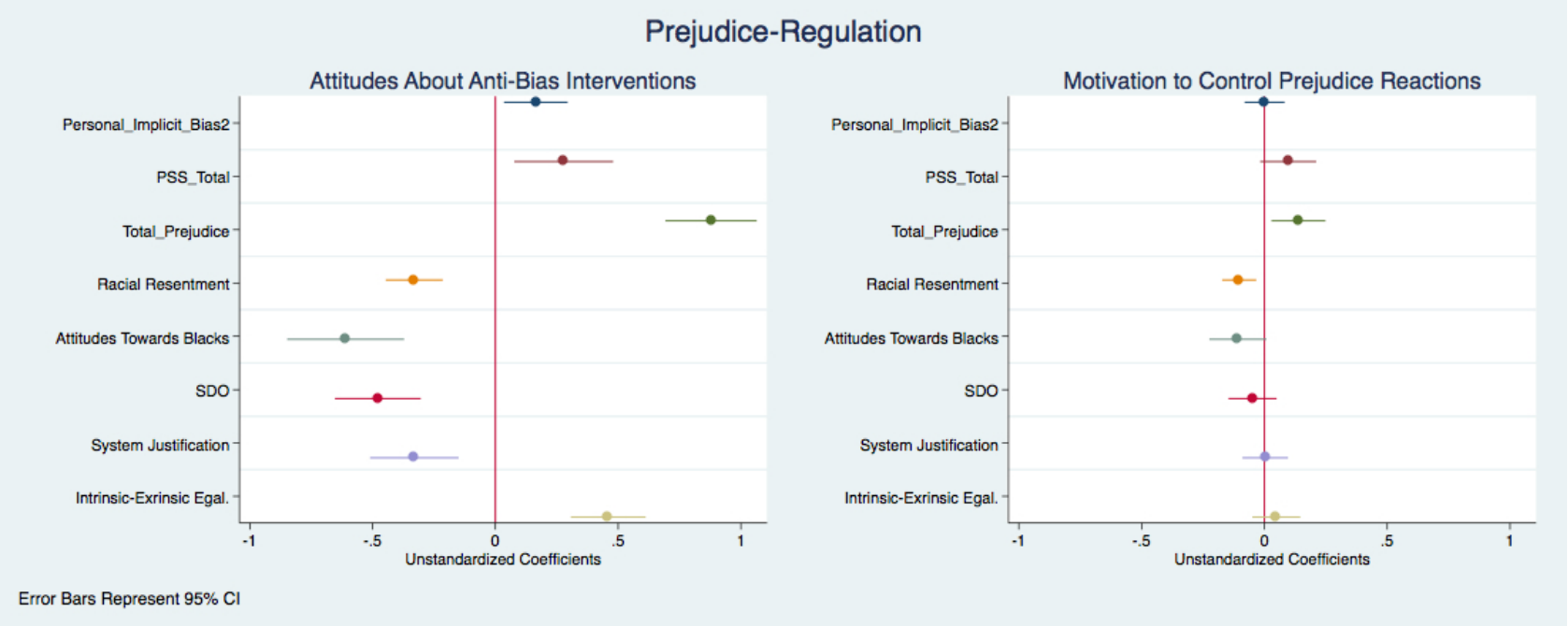


Figure 1.5

Coefficients plot for effect of T1 independent variables on stereotyping and public policy attitudes in Study 1

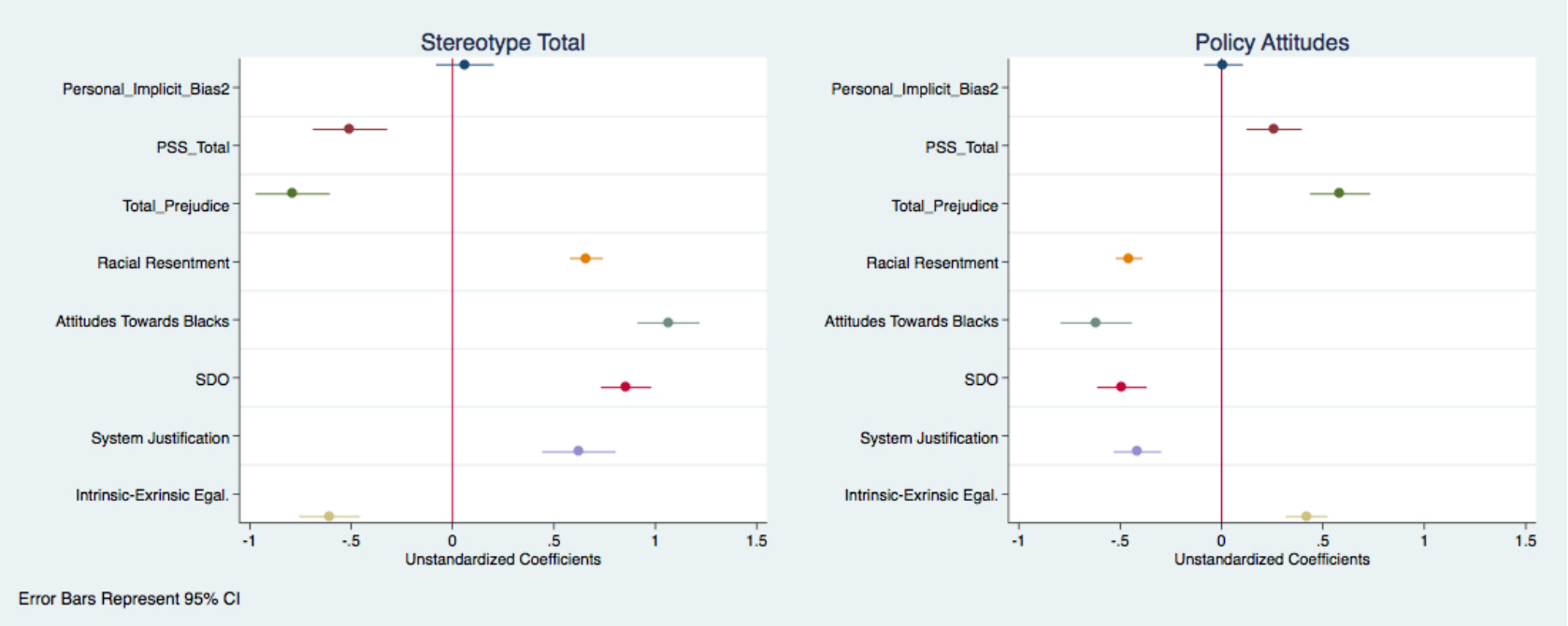


Figure 1.6

Experimental condition x social dominance orientation, system justification, and explicit black attitudes on stereotype endorsement.

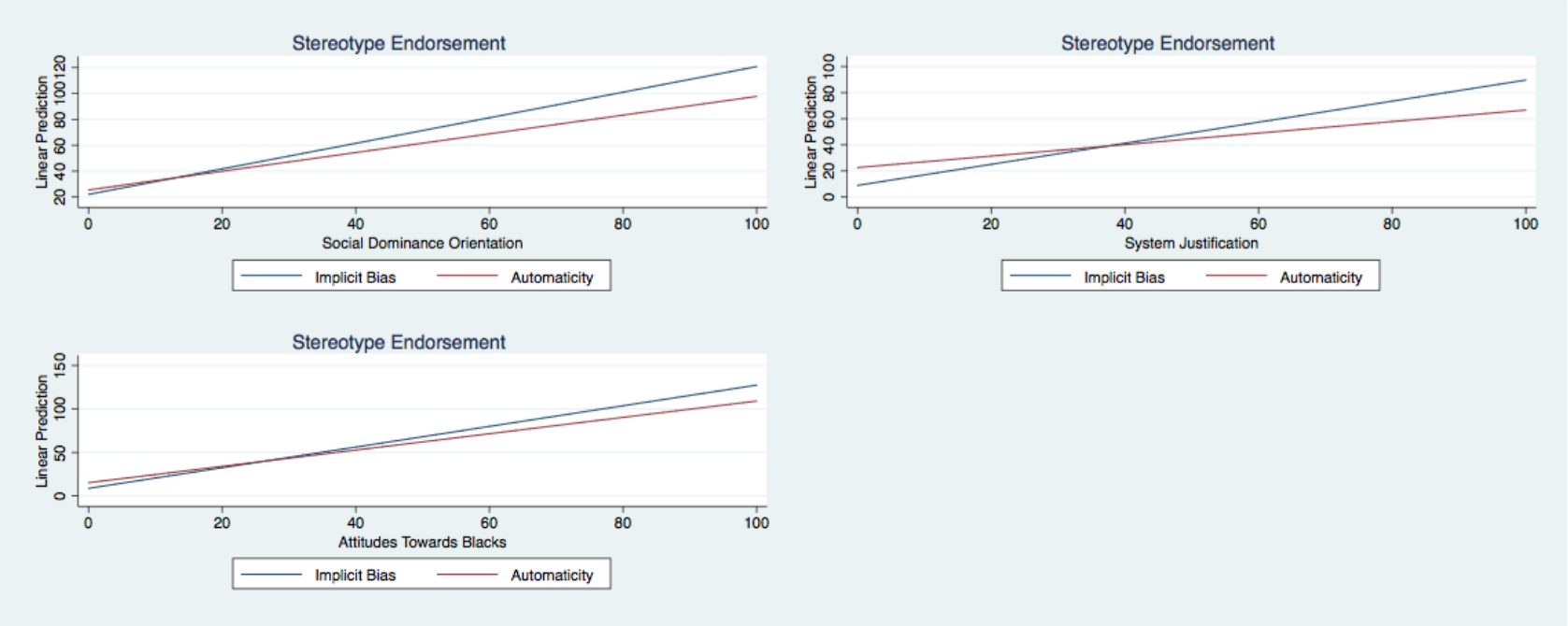


Figure 2.1
Main effect of experimental condition on affect for Pilot 2

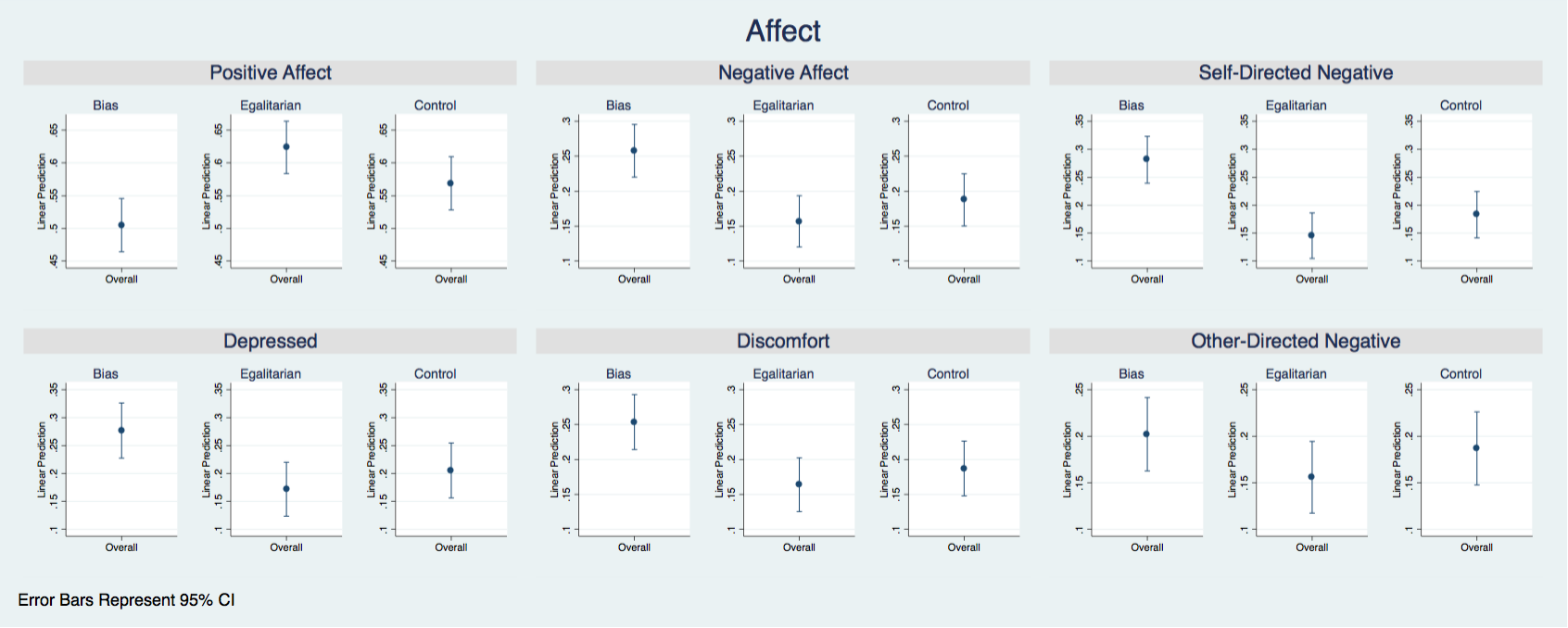


Figure 2.1
Main effect of experimental condition on affect for Pilot 2

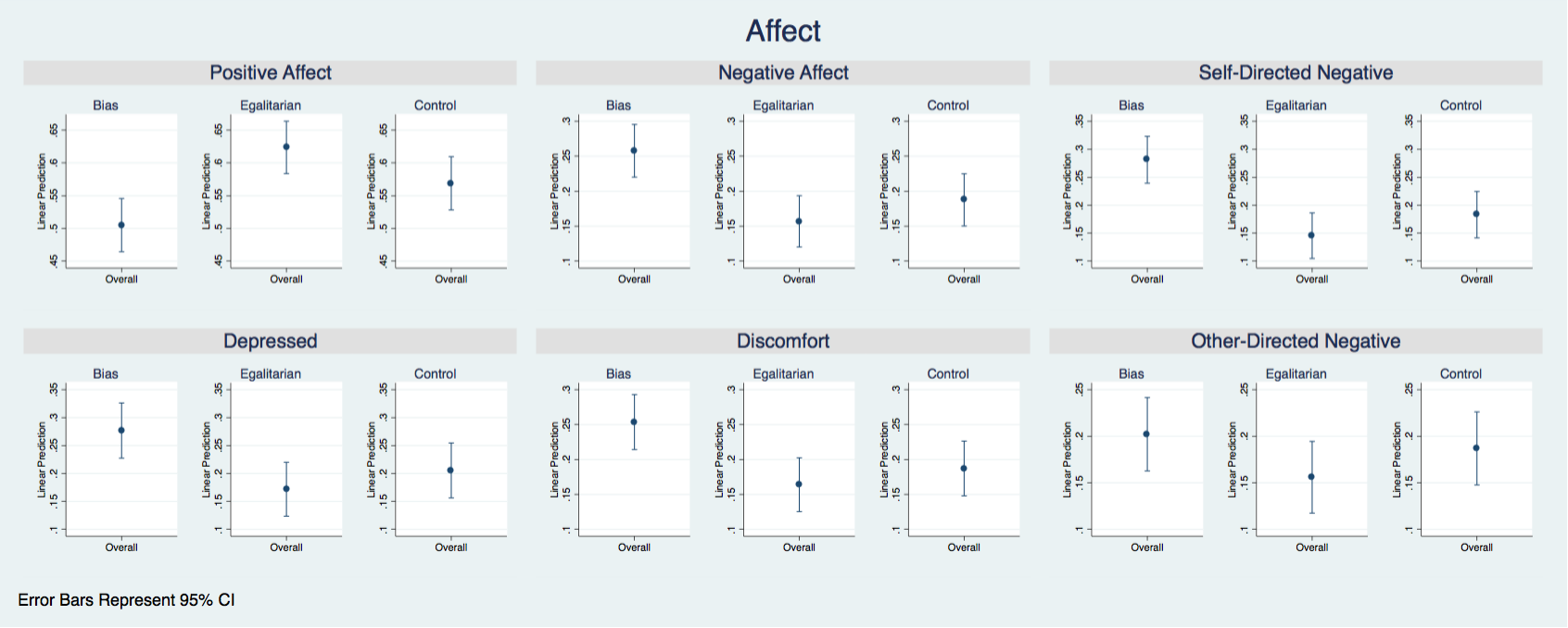


Figure 2.2
Main effect of experimental condition on motivated reasoning for Pilot 2

Figure 2.3
Coefficients plot for effect of affect on motivated reasoning in Pilot 2

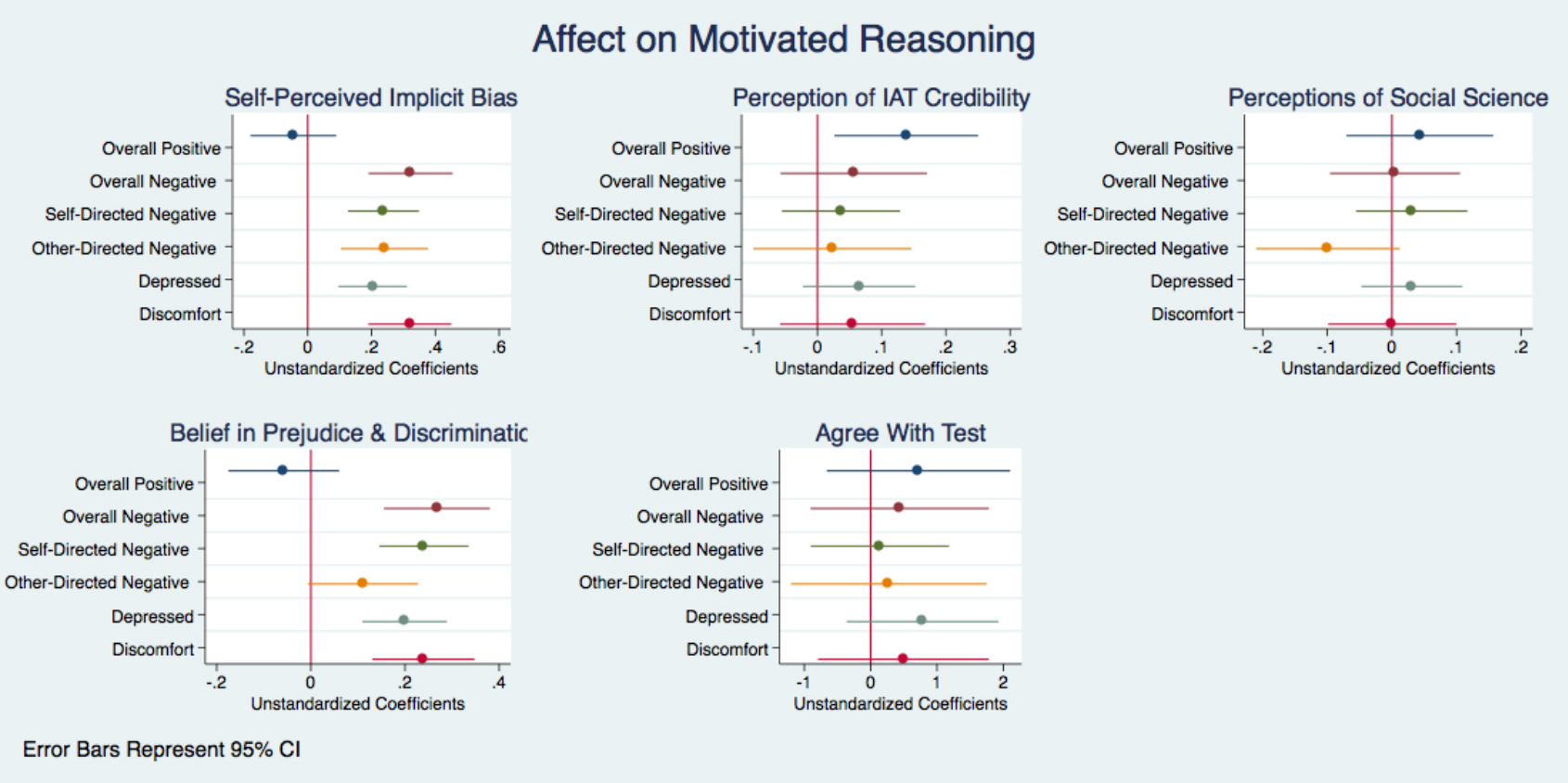


Figure 2.4
Main effect of experimental condition on affect for Study 2

Error Bars Represent 95% CI

Figure 2.5
SDO x experimental condition on affect for Study 2

Figure 2.6
Coefficients plot for effect of affect on motivated reasoning in Study 2

Figure 2.7

Main effect of experimental condition on motivated reasoning for Study 2

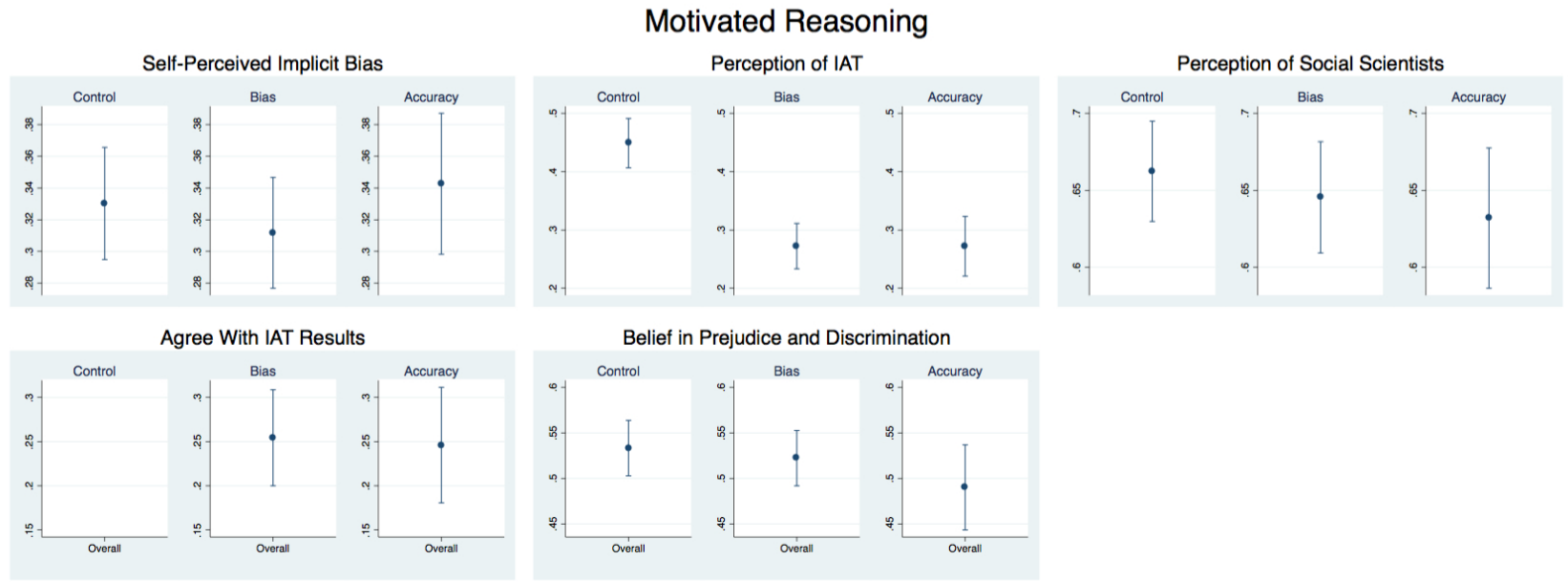


Figure 2.8
Coefficients plot for effect of individual differences on motivated reasoning in Study 2

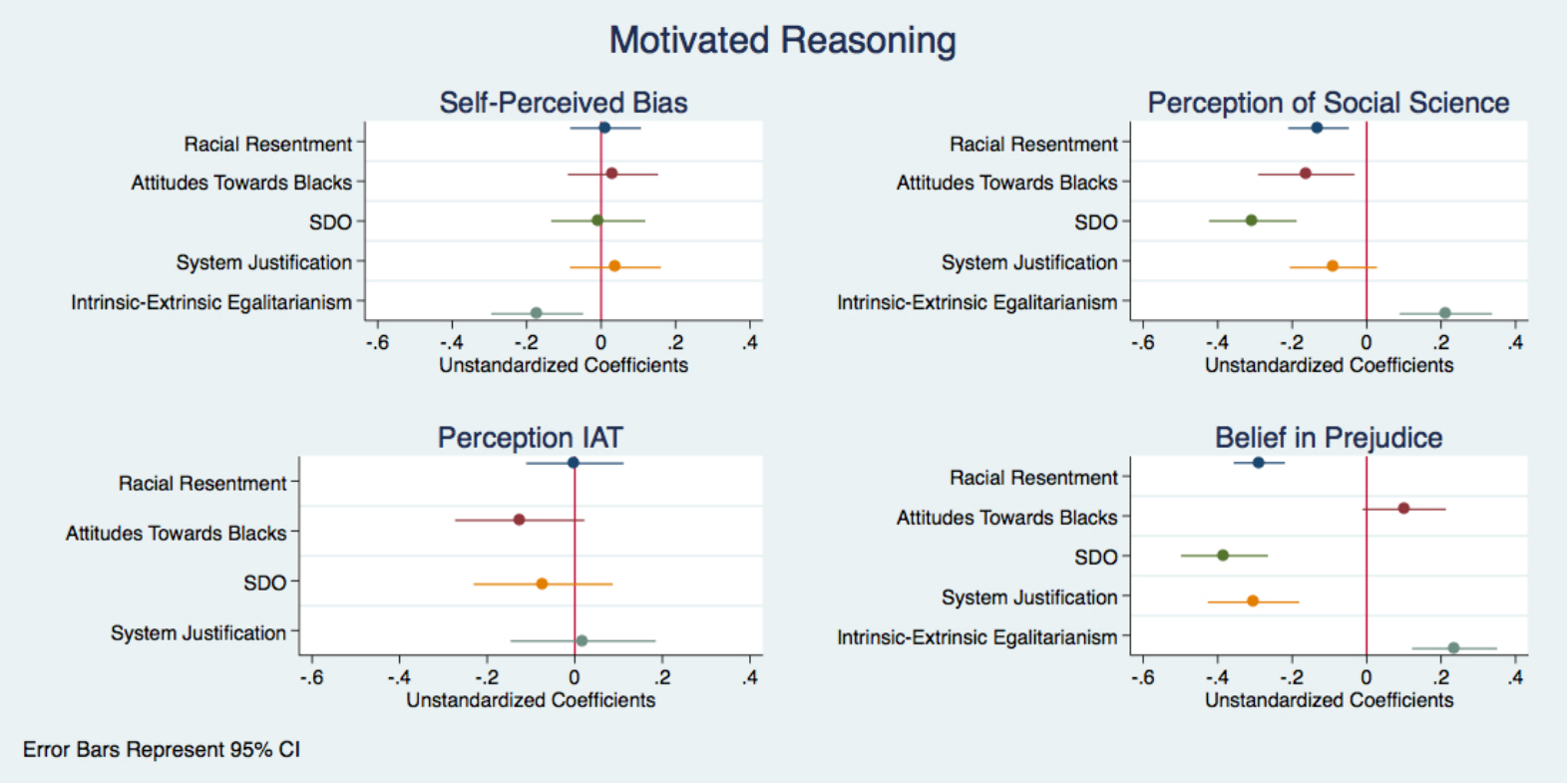


Figure 2.9
SDO x experimental condition on belief in prejudice and discrimination for Study 2

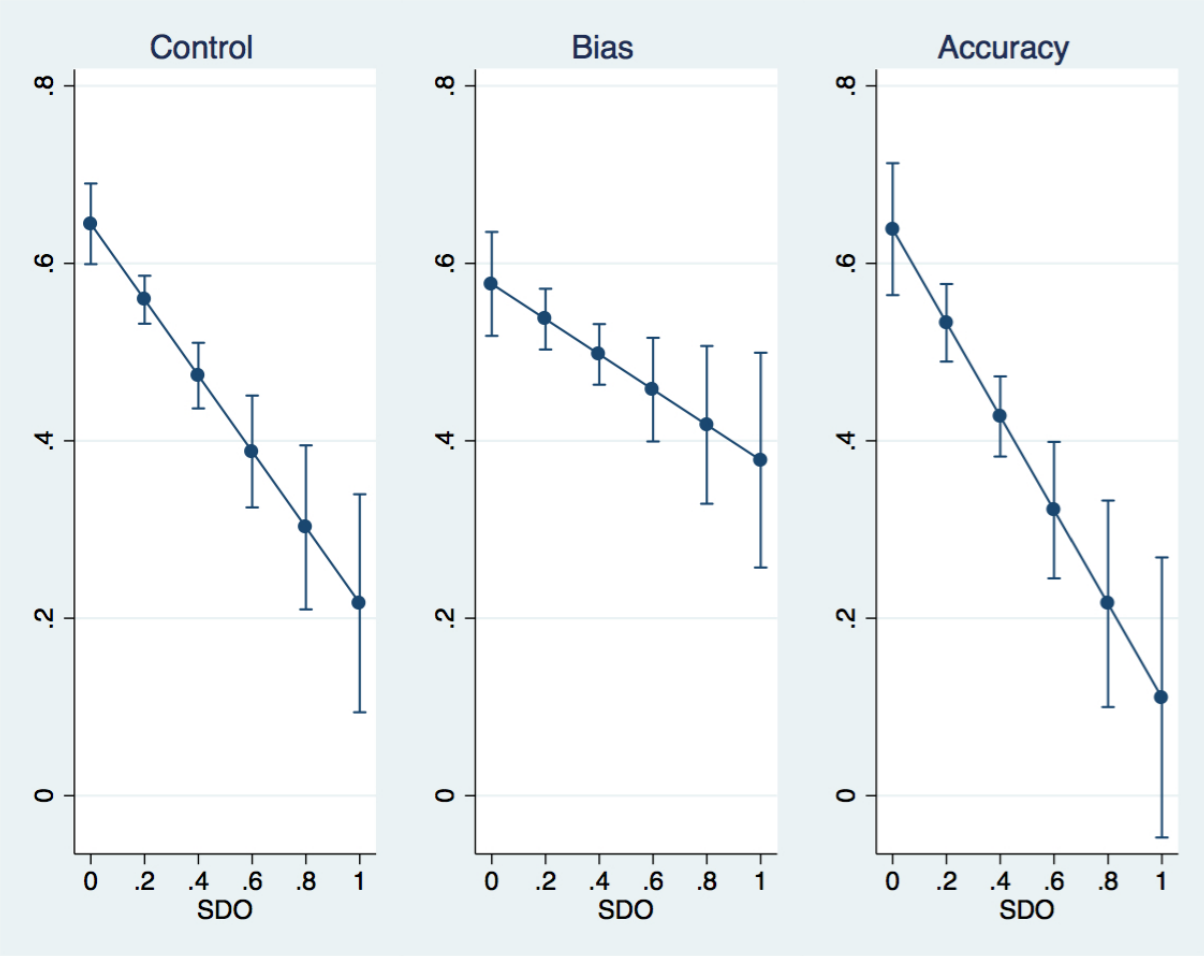


Figure 2.10
Egalitarianism x experimental condition on perception of IAT for Study 2

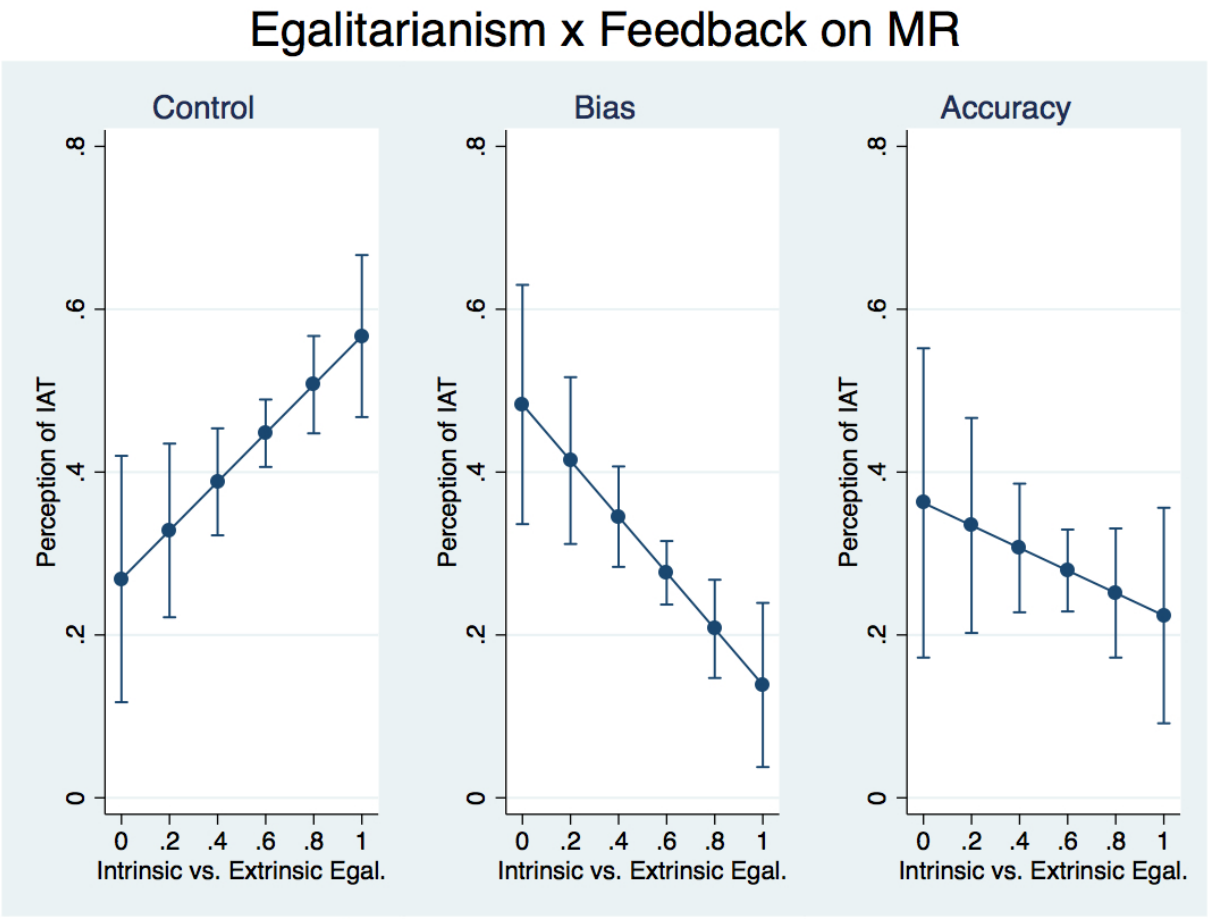


Figure 2.11
Egalitarianism x experimental condition on belief in prejudice and discrimination for Study 2

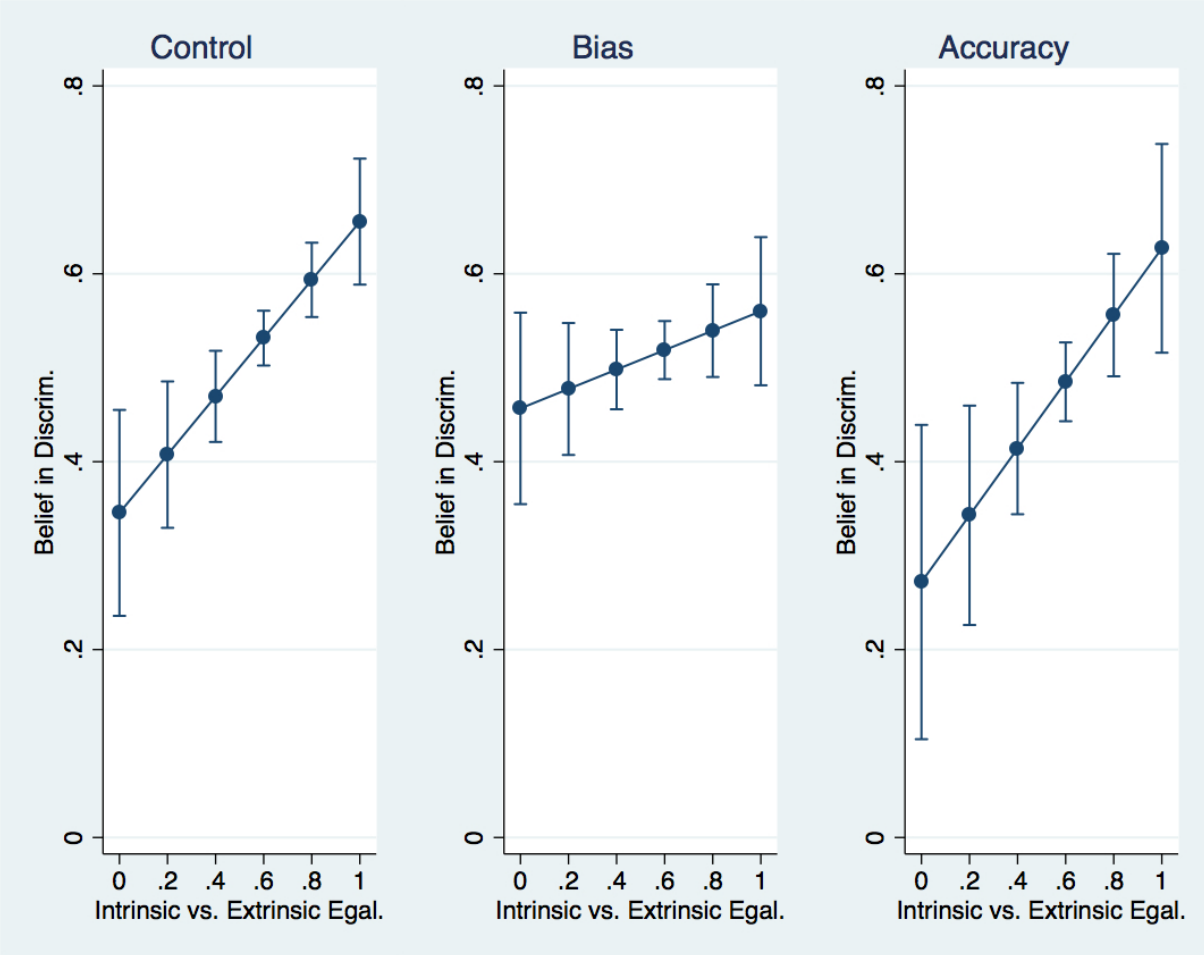


Figure 2.12
The relationship between attitudes towards blacks and attitudes towards anti-bias interventions, separated by experimental condition

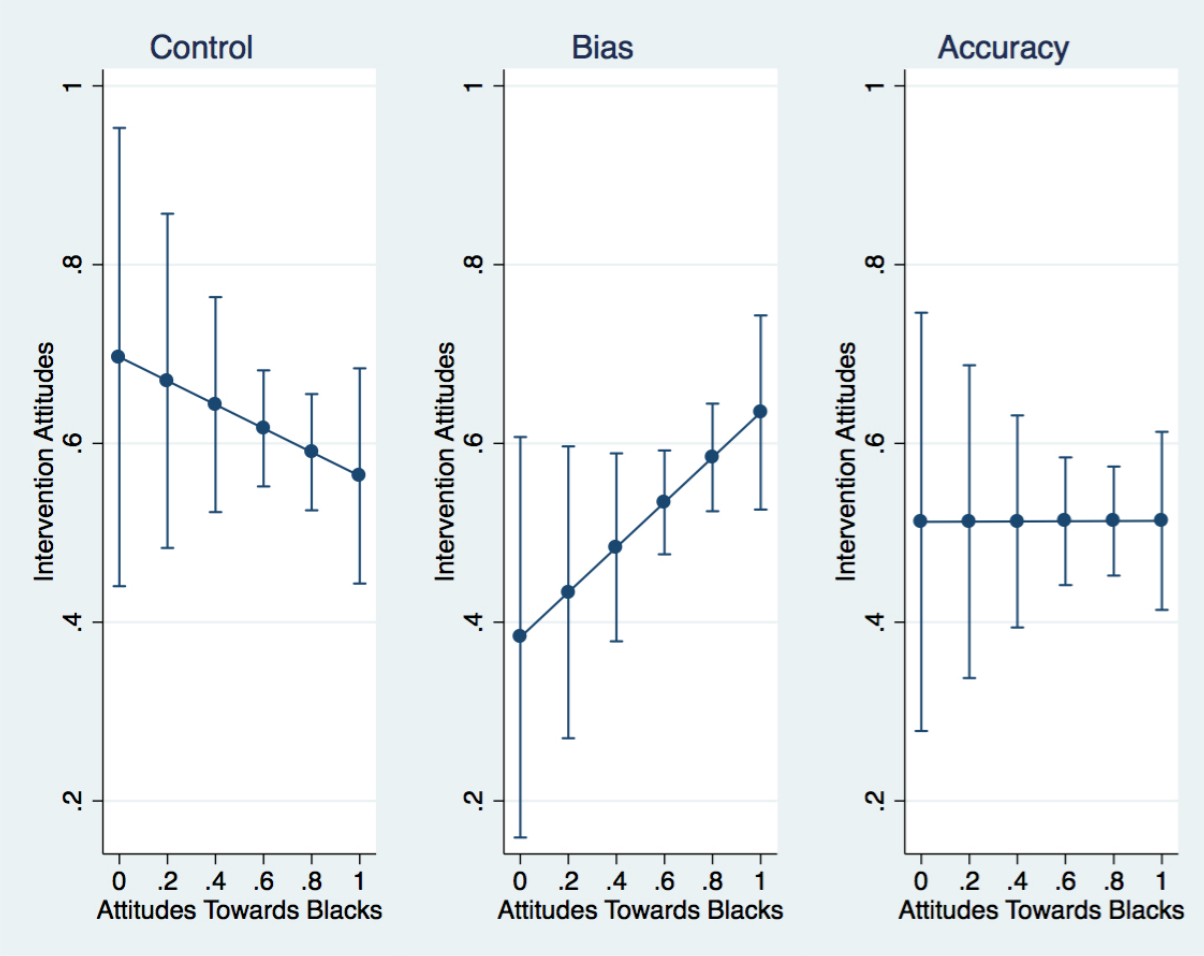


Figure 2.13

Coefficients plot for effect of individual differences on prejudice-regulation in Study 2

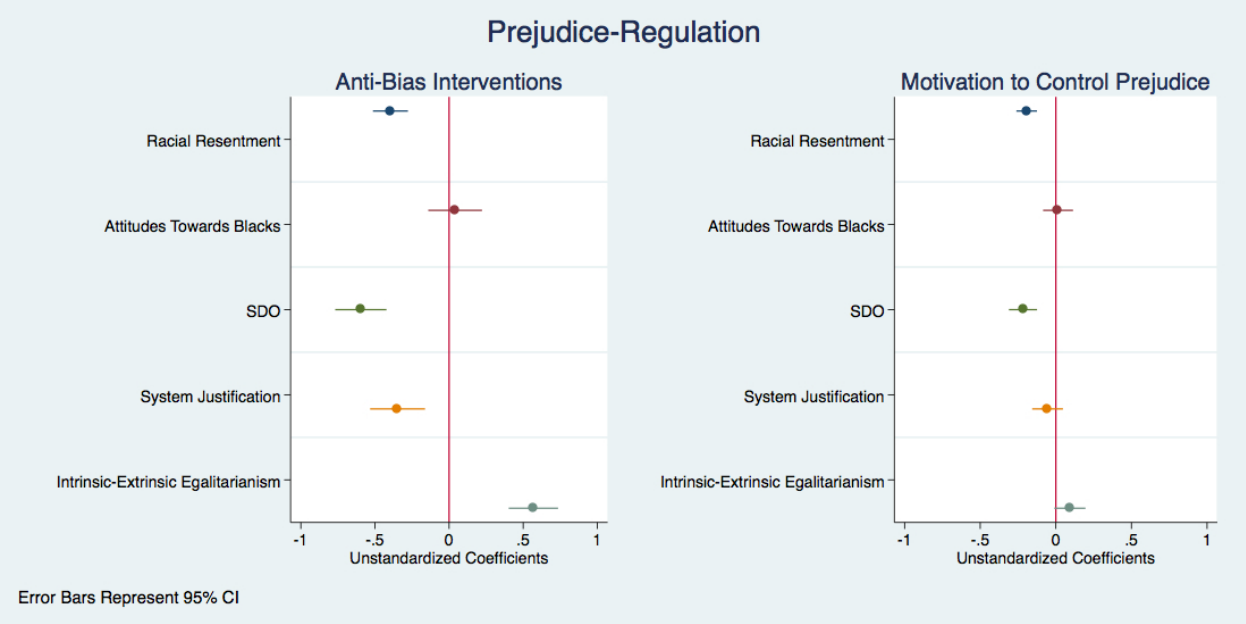


Figure 2.14

Coefficients plot for effect of individual differences on stereotyping and policy attitudes in Study 2

Figure 2.15

The relationship between system justification and attitudes towards anti-bias interventions, separated by experimental condition

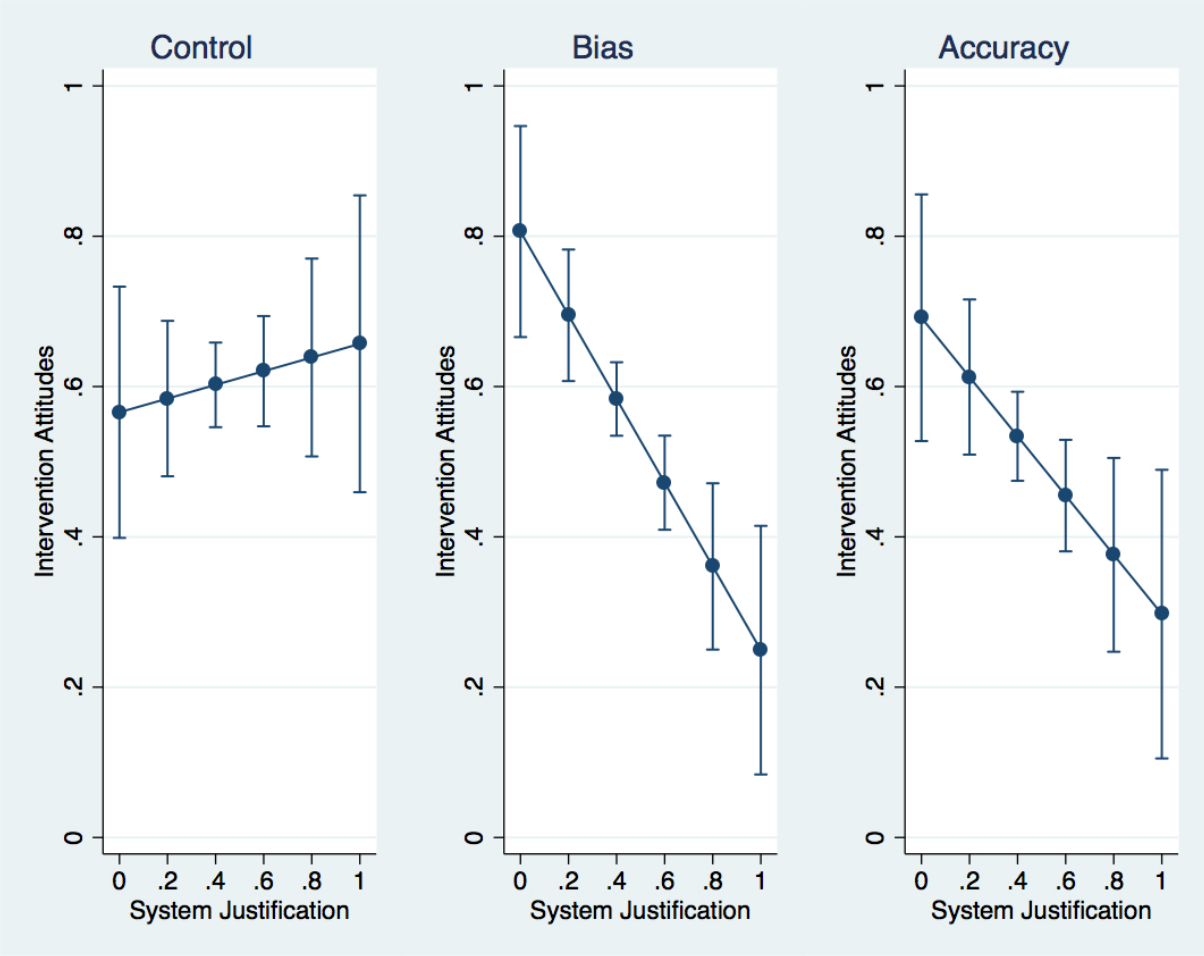


Figure 3.1

Coefficients plot for effect of affect on motivated reasoning in Pilot 3

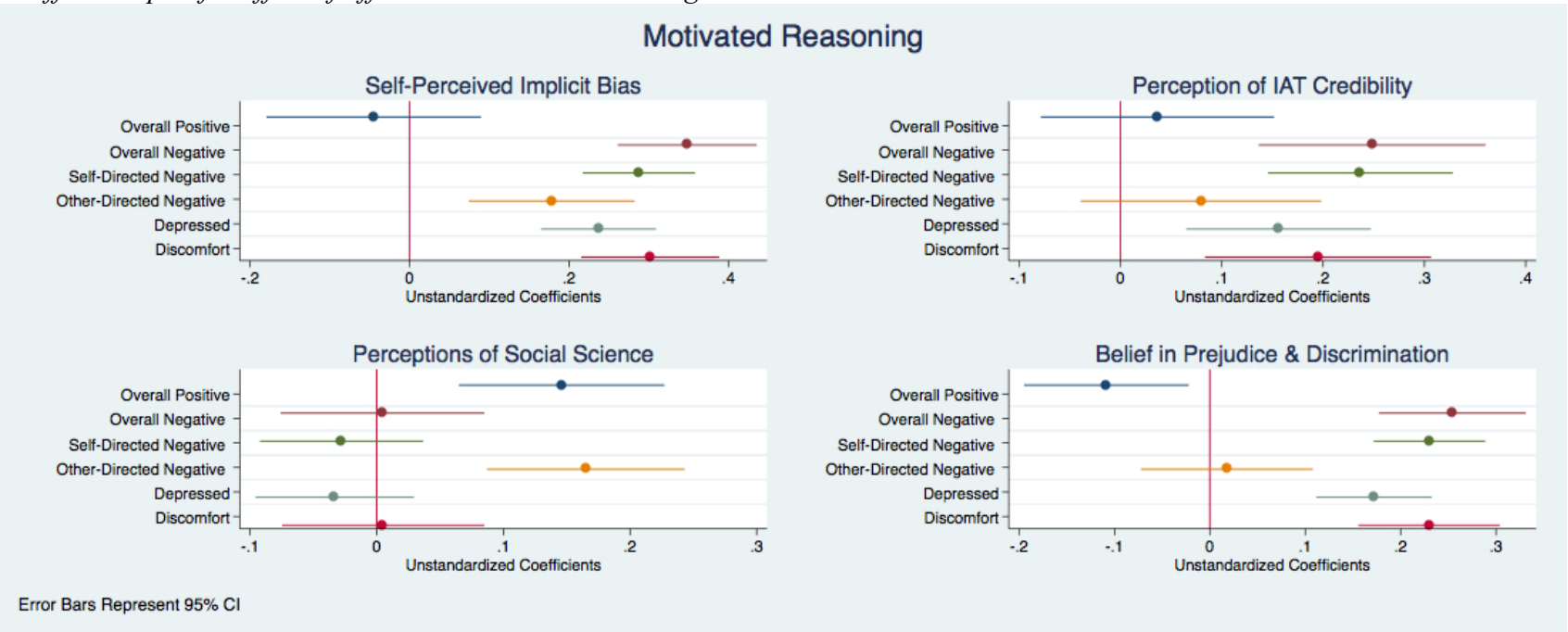


Figure 3.2
Main effect of experimental condition on affect for Study 3a

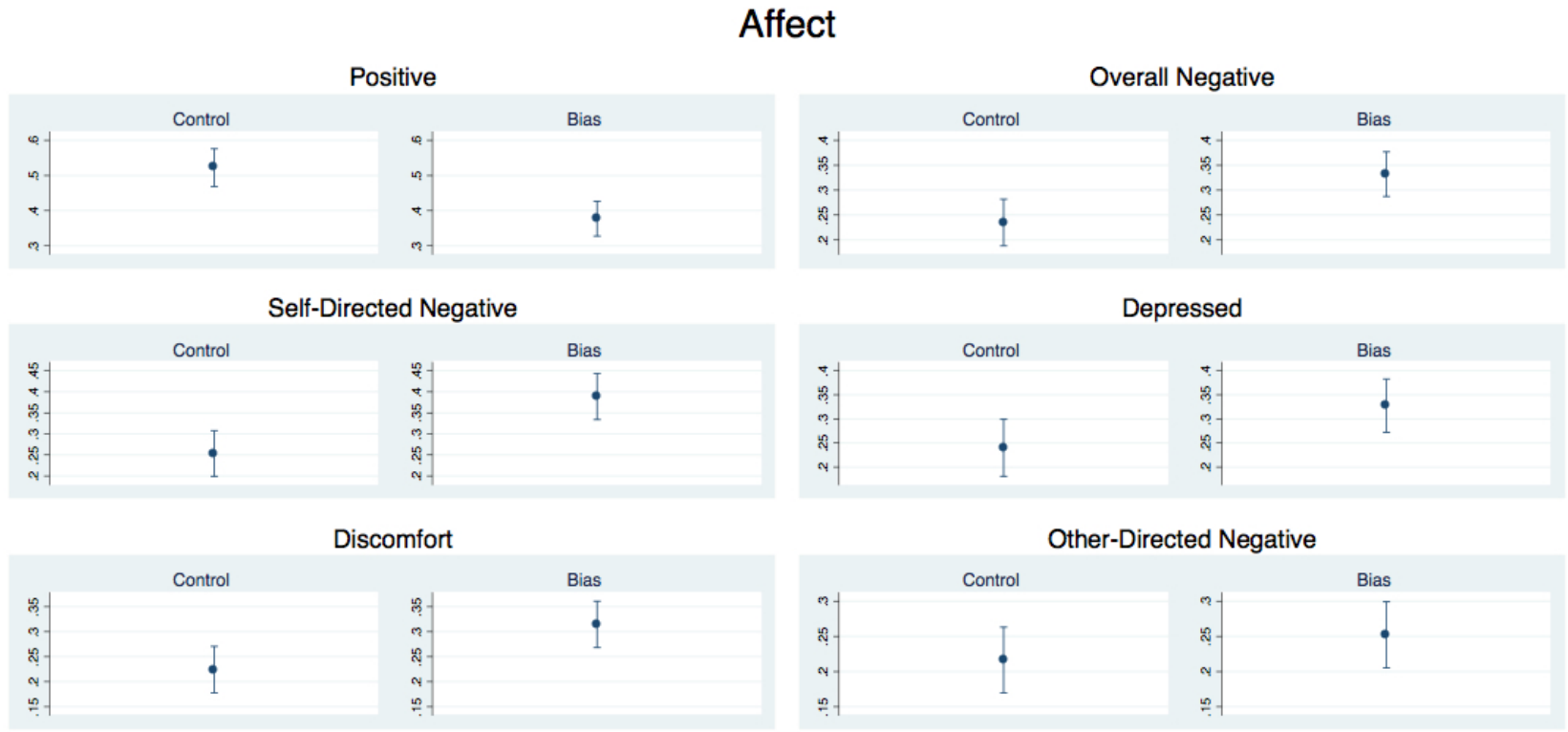


Figure 3.3
Main effect of experimental condition on affect for Study 3a

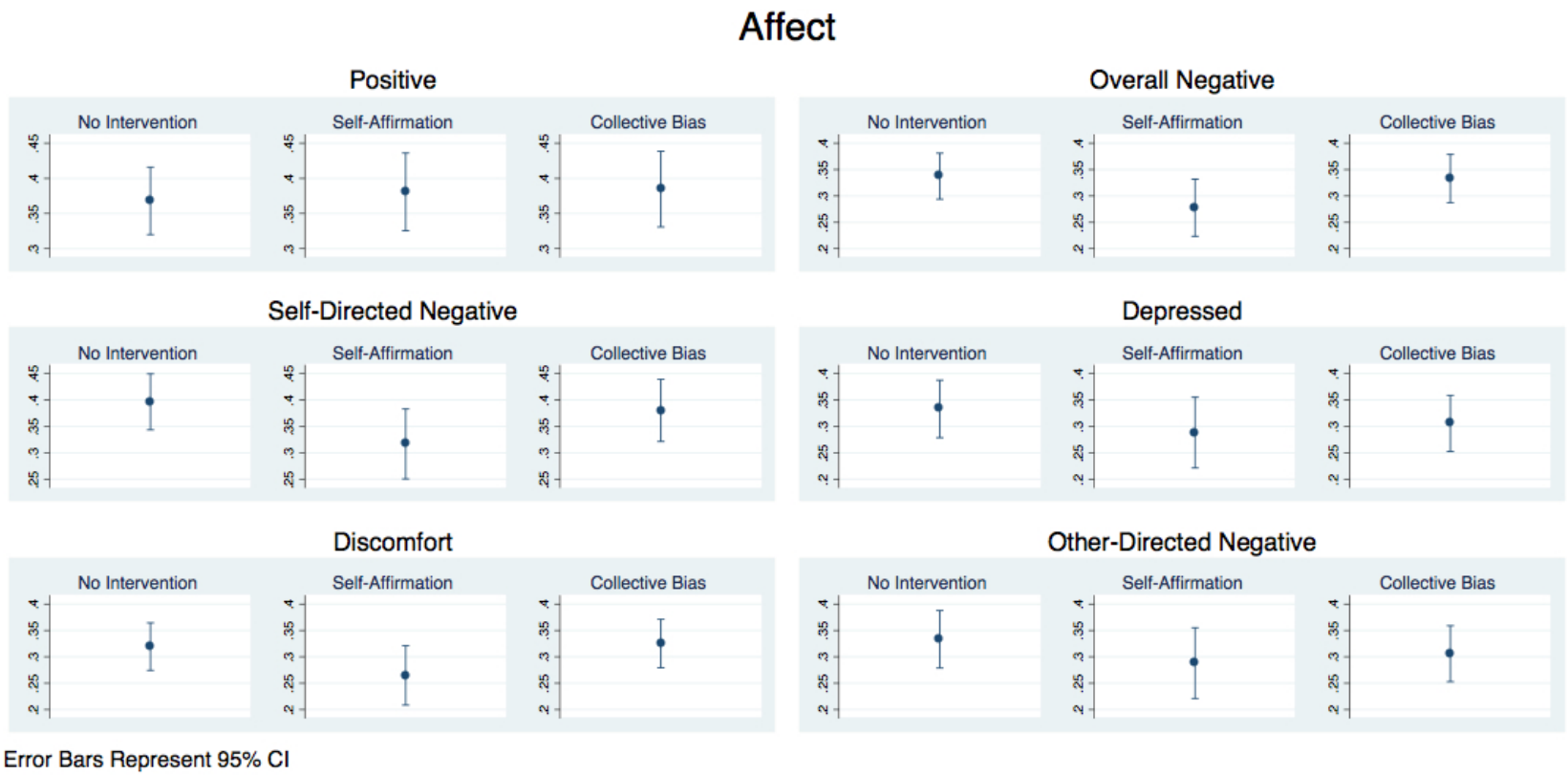
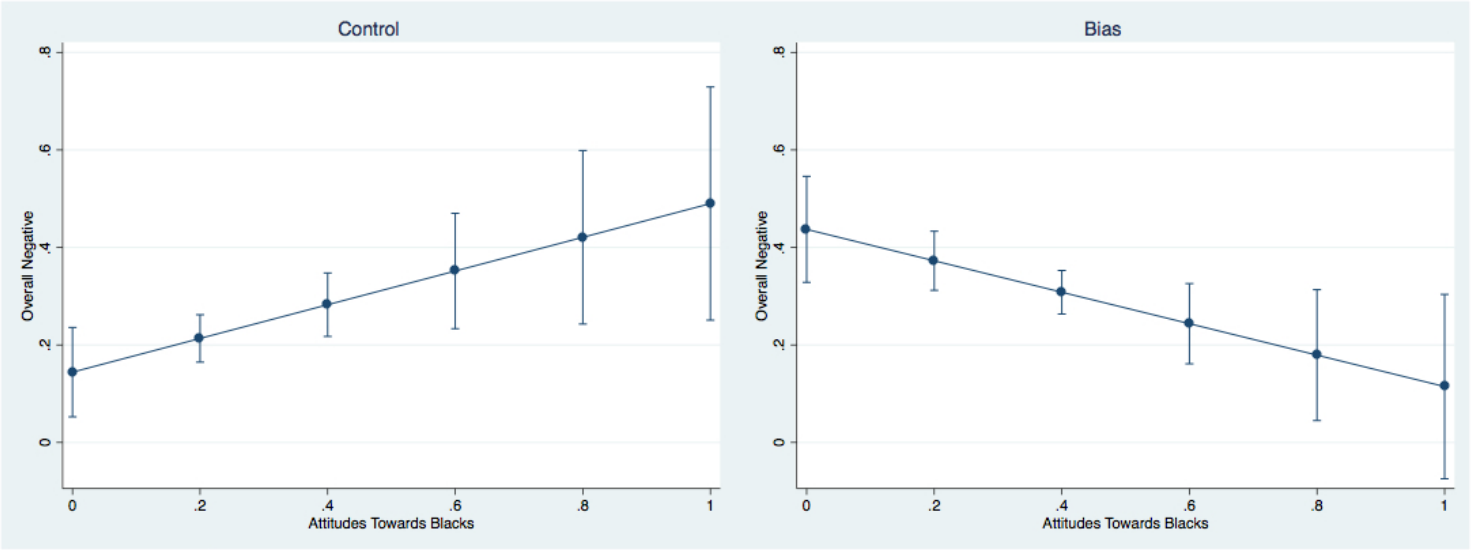


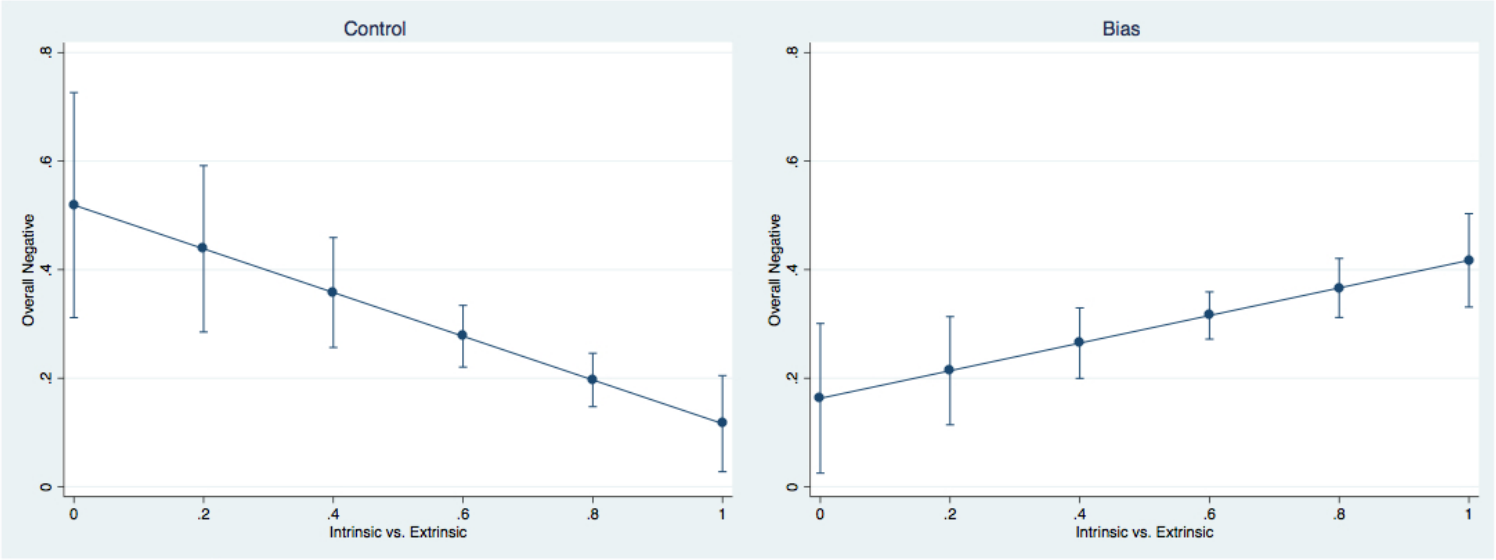
Figure 3.4

Attitudes towards blacks x experimental condition on affect for Study 3a



Error Bars Represent 95% CI

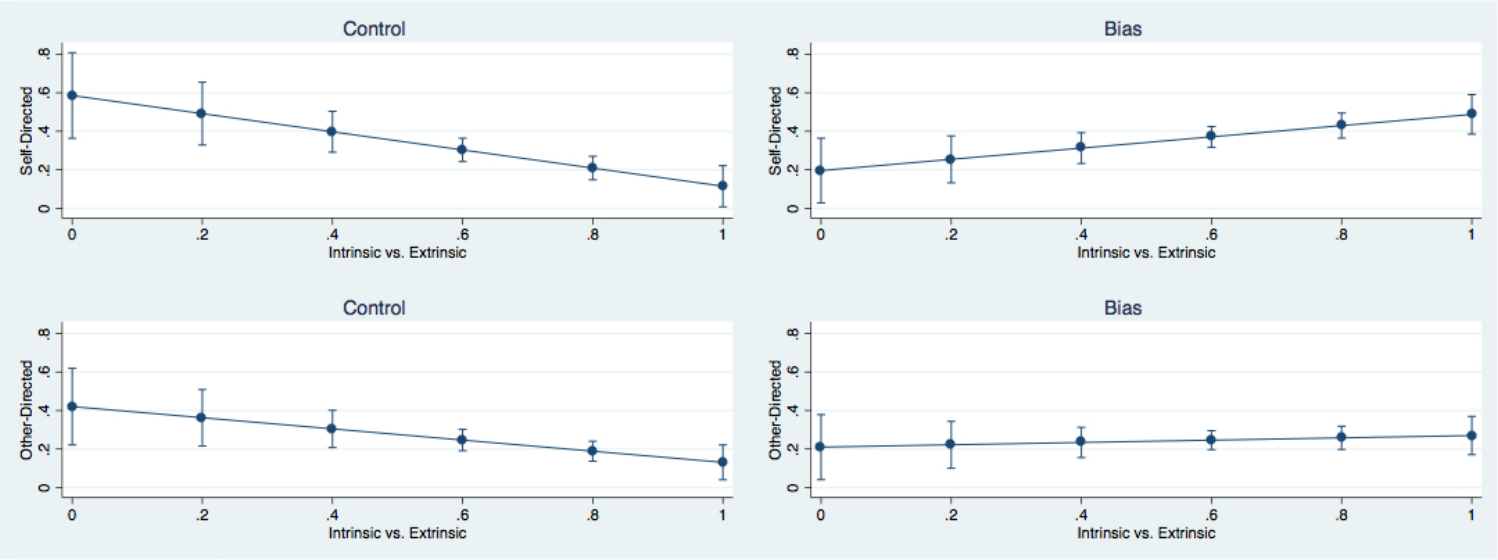
Figure 3.5
Intrinsic-extrinsic egalitarianism x experimental condition on negative affect for Study 3a



Error Bars Represent 95% CI

Figure 3.6

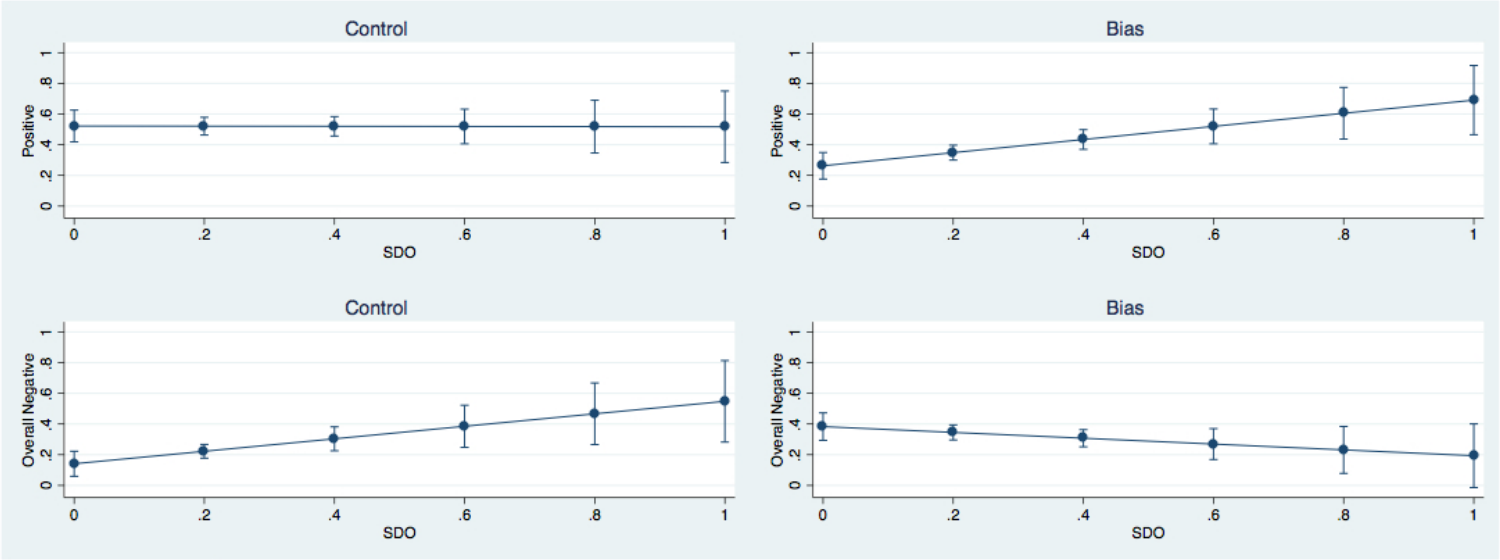
Intrinsic-extrinsic egalitarianism x experimental condition on self- and other-directed negative affect for Study 3a



Error Bars Represent 95% CI

Figure 3.7

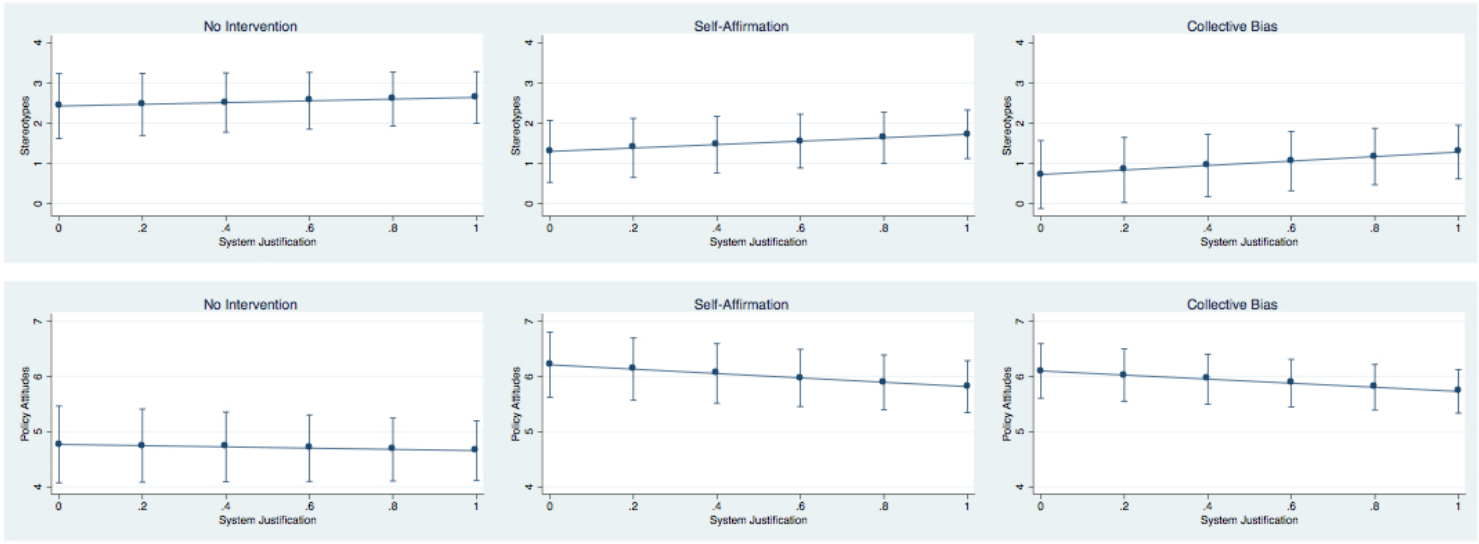
SDO x experimental condition on self- and other-directed negative affect for Study 3a



Error Bars Represent 95% CI

Figure 3.8

SJ x intervention condition on self-perceived bias, stereotyping, and policy attitudes for Study 3a



Error Bars Represent 95% CI

Figure 3.9
Main effect of experimental condition on affect for Study 3b

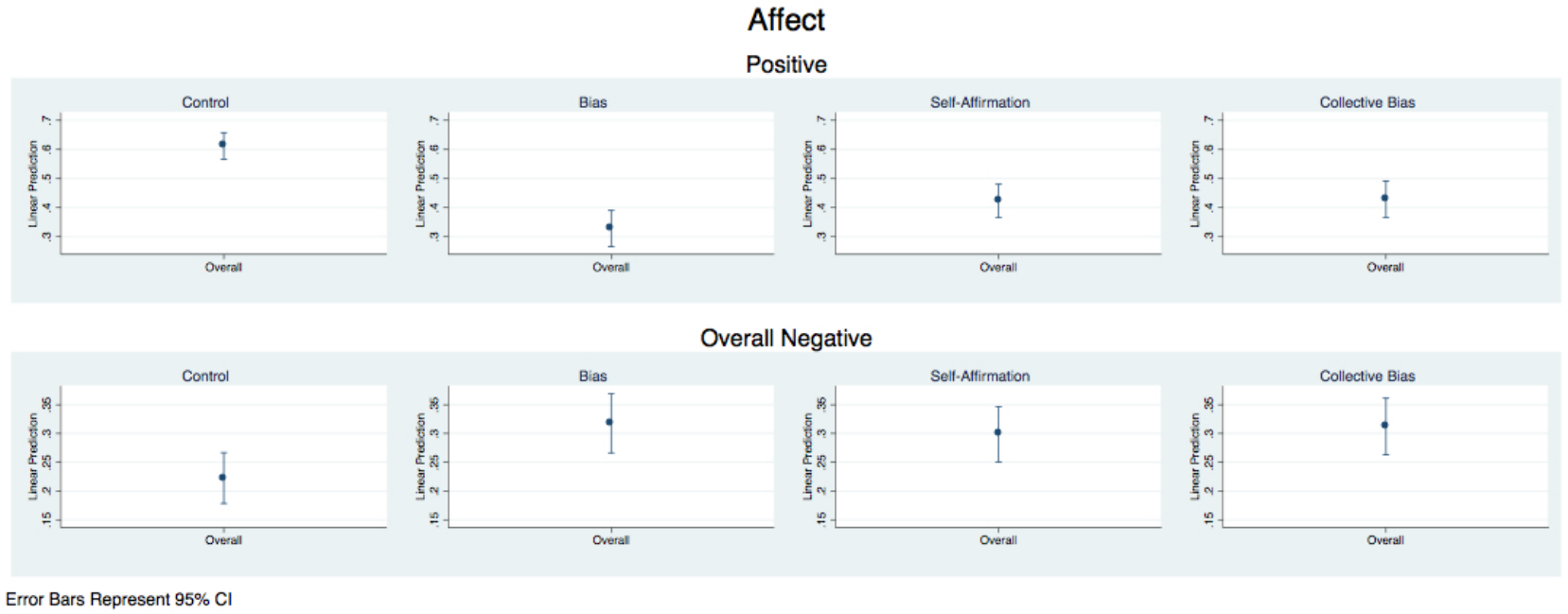


Figure 3.10
Main effect of experimental condition on Motivated Reasoning for Study 3b

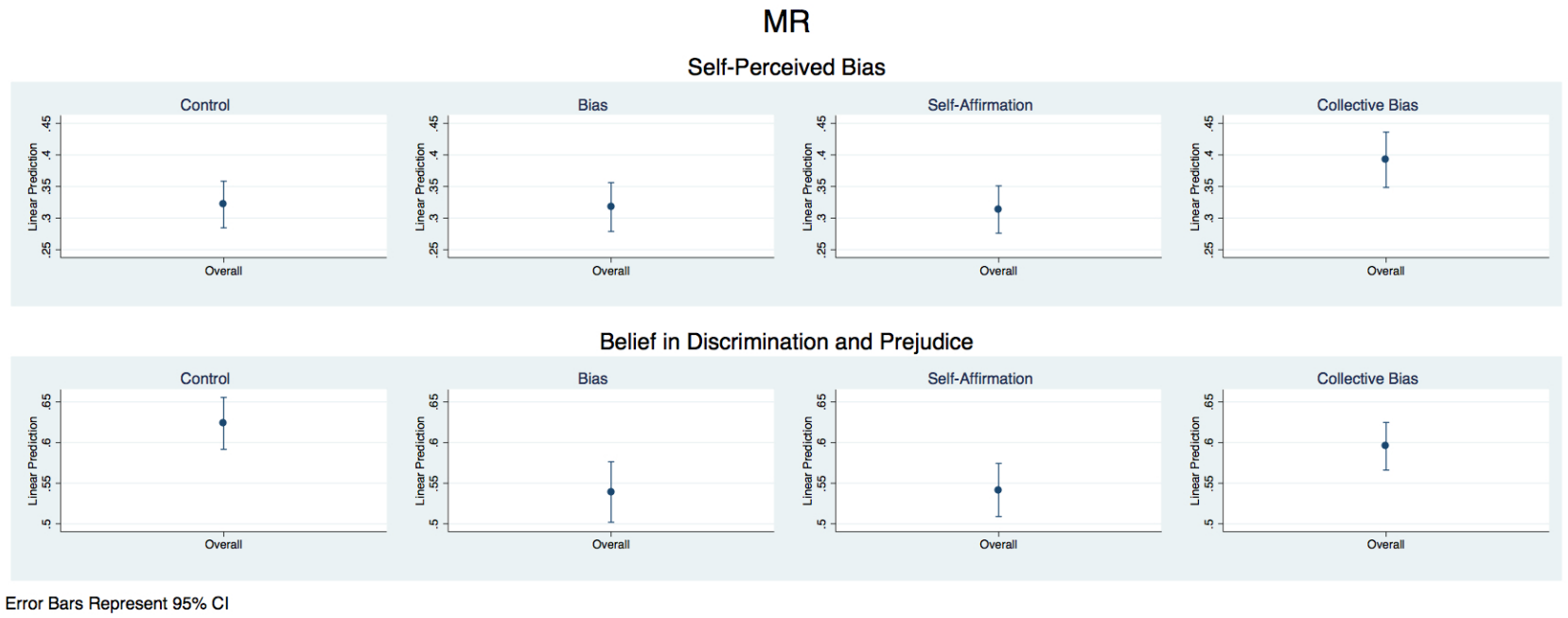
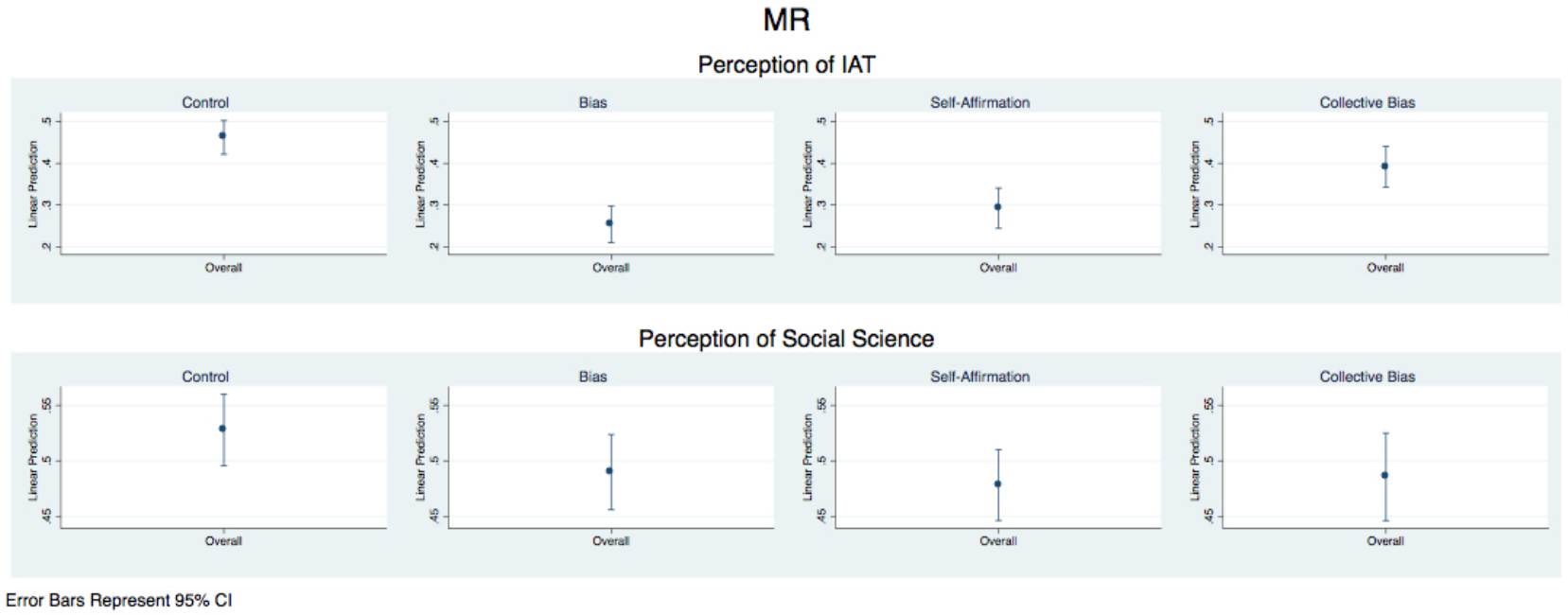


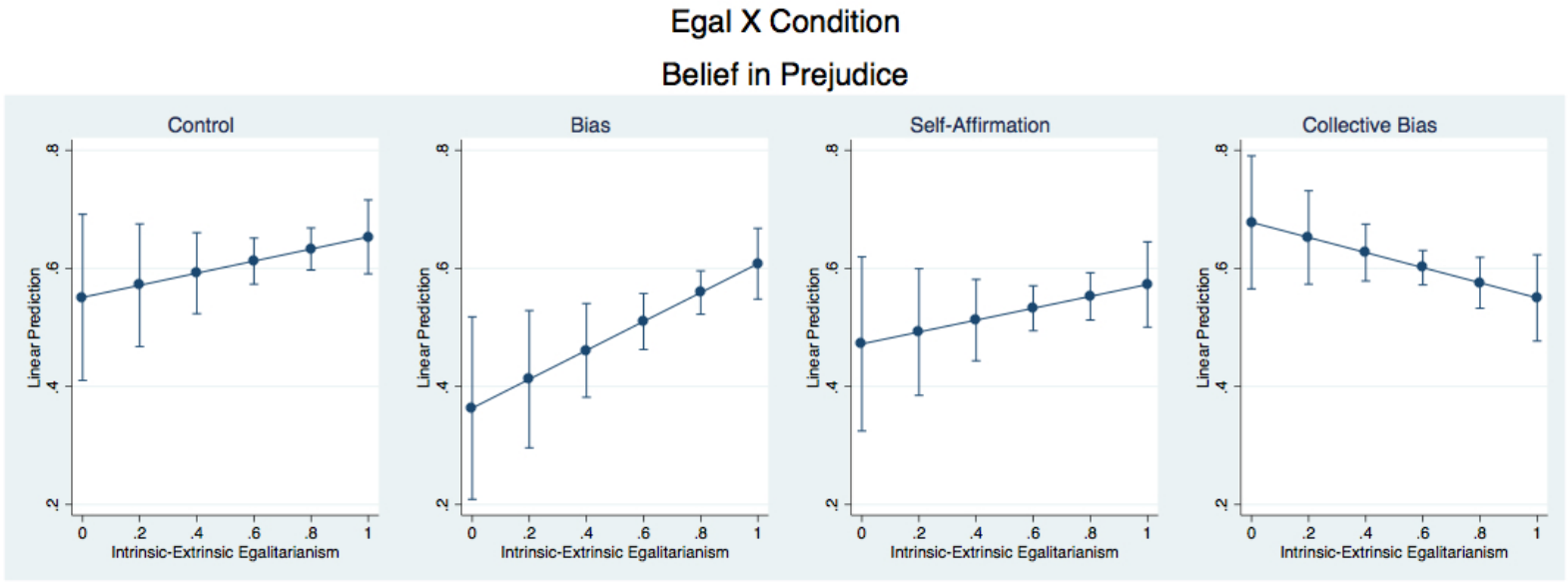
Figure 3.11
Main effect of experimental condition on Motivated Reasoning for Study 3b



Error Bars Represent 95% CI

Figure 3.12

Egalitarianism x experimental condition on belief in prejudice for Study 3b



Error Bars Represent 95% CI

Figure 3.13
SDO x experimental condition on belief in prejudice for Study 3b

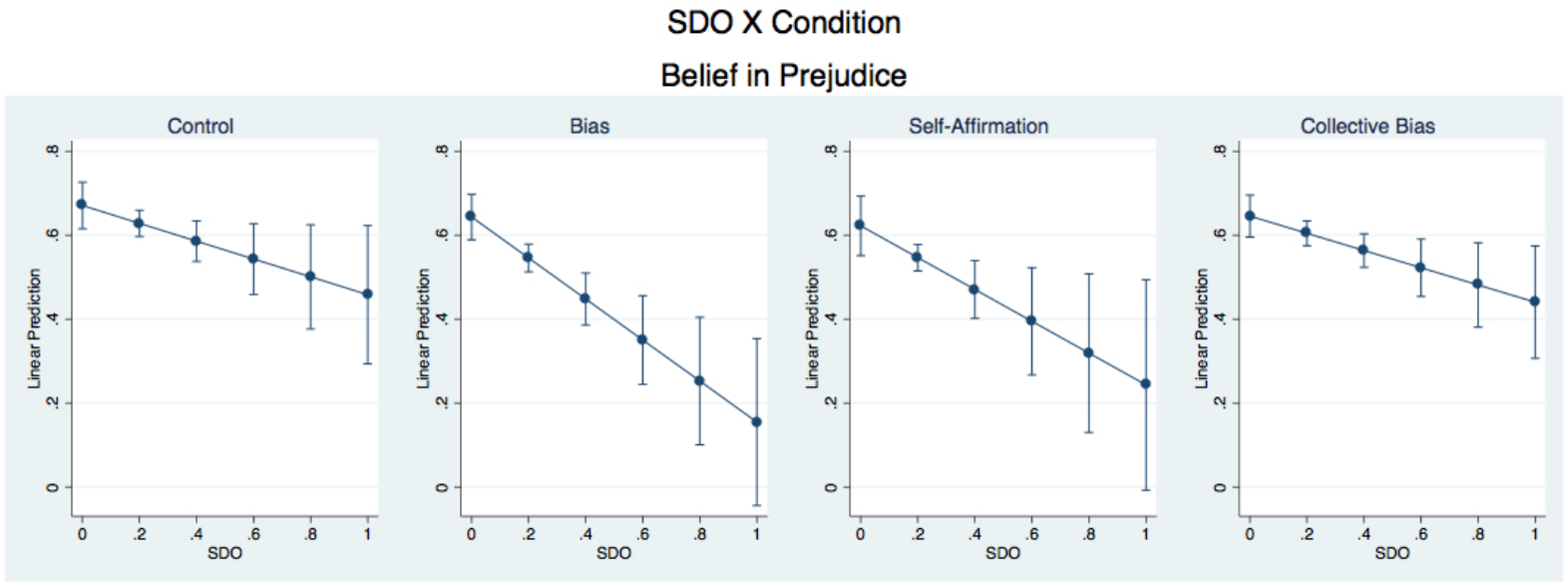


Figure 3.14
SDO x experimental condition on perception of IAT for Study 3b

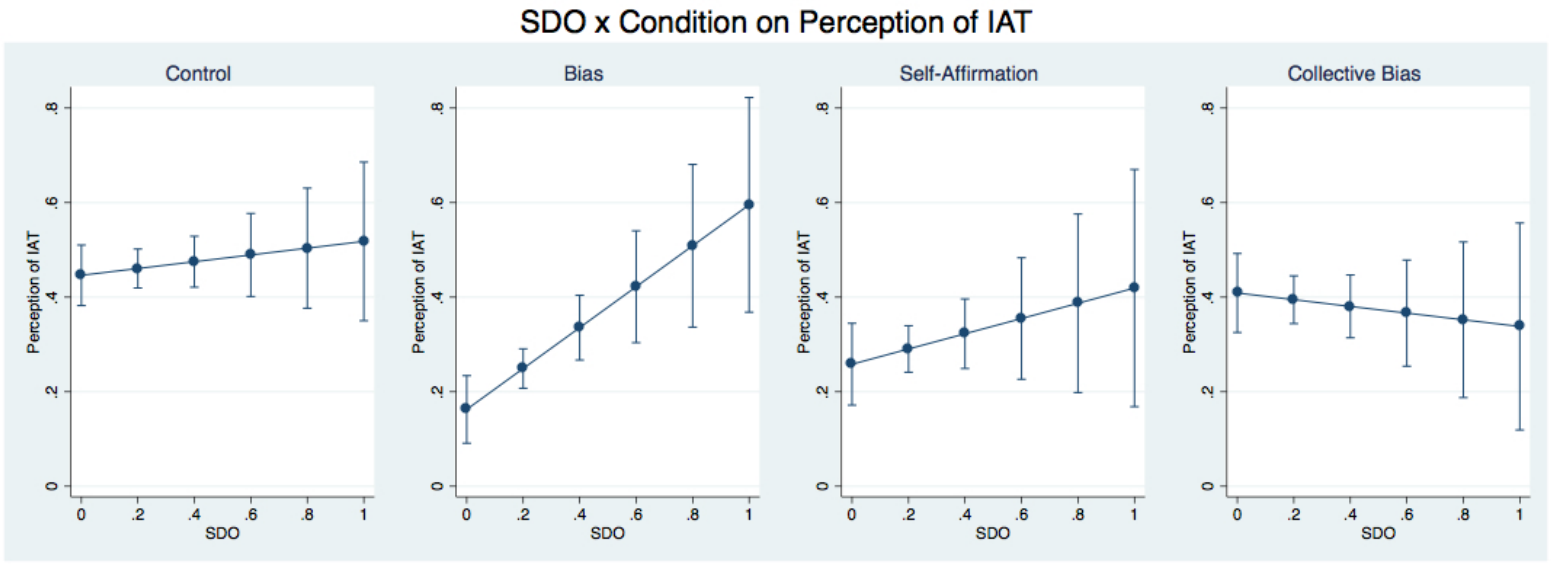
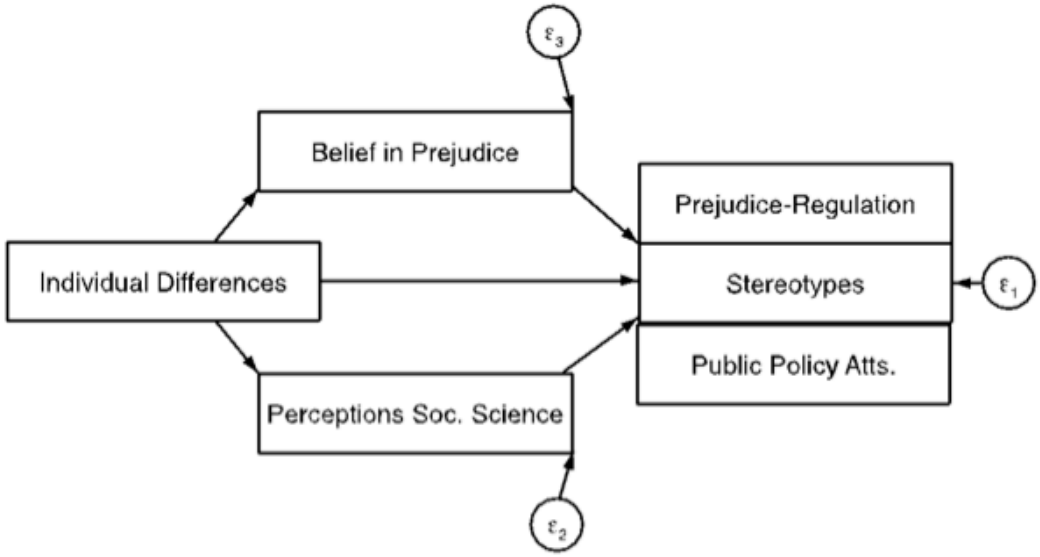


Figure A. Conceptual model for motivated reasoning processes as a mediator for the relationship between individual differences and prejudice-regulation, stereotyping, and public policy attitudes.



Appendix

Hypotheses

1. People characterized by explicitly hostile racial attitudes, resentment towards racial progress, system justification or social dominance orientation will be motivated to reject all evidence of aggregate-level racial bias and discrimination, as well as evidence of their own implicit racial bias (Hypothesis 1).
2. Evidence of aggregate-level racial prejudice and discrimination, and feedback that one harbors implicit racial bias, will strengthen or activate the effect of racial resentment, system justification, and social dominance orientation on the motivated rejection of all evidence of racial discrimination and inequality and implicit racial bias (Hypothesis 2)
3. Acceptance of the existence of implicit bias in the general population among intrinsic egalitarians should be particularly likely when these individuals encounter credible evidence to support that inference. However, people characterized by an extrinsic motivation to be egalitarian will be unwilling to acknowledge the general existence of implicit racial bias. (Hypothesis 3)
4. Two complementary predictions:
    a. Feelings of guilt and shame among low prejudiced, intrinsic egalitarians will attenuate the motivated rejection of evidence of their own implicit racial bias and increase prejudice regulation (Hypothesis 4a**).**
    b. Extrinsic egalitarians will be fundamentally threatened by, avoidant of, and motivated to reject evidence of their own unconscious racial bias (Hypothesis 4b).
5. Intrinsic (vs. extrinsic) egalitarians with dissociated implicit-explicit racial attitudes (e.g., low explicit prejudice, high implicit prejudice) will be more accepting of evidence of their own implicit racial bias (Hypothesis 5)
6. High prejudiced individuals who are motivated to project an egalitarian self-image primarily for extrinsic reasons or in compliance with socially imposed norms and expectations to be motivated to reject evidence of their own unconscious racial bias (Hypothesis 6)
7. The motivated rejection of implicit bias evidence is expected to increase personal (vs. situational) attributions for the causes of aggregate-level racial disparities (Hypothesis 7)
8. Personal (vs. situational) attributions for the causes of aggregate-level racial disparities will increase support for more punitive attitudes towards the criminal justice system (Hypothesis 8).
9. Personal (vs. situational) attributions for the causes of aggregate-level racial disparities will increase opposition to policy remediation to address racial inequality such as affirmative action and social welfare (Hypothesis 9).
10. Effects of motivated reasoning processes will undermine individuals' perceived value and importance of organizational interventions designed to reduce implicit racial bias (Hypothesis 10).

11. Effects of motivated reasoning processes will reduce the general motivation to control prejudiced responding and its impact on behavior (Hypothesis 11; Dunton & Fazio, 1997).
12. Two independent interventions will reduce the motivated rejection of allegations of one's own implicit racial bias and its downstream effects described above (Hypothesis 12).

Measures Index

Study 1 Pre-Manipulation (T1)

1) Explicit Racial Attitudes
    a. Racial Resentment (Kinder & Sanders, 1996)
    b. Attitudes Toward Blacks Scale (Brigham, 1993)
2) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
3) Authoritarianism (Stenner, 2005)
4) Internal/External Motivation to Control Prejudice (Plant & Devine, 1998)
5) Social Dominance Orientation (Sidanius & Pratto, 2001)
6) System justification (Kay & Jost, 2003)
7) Stereotype threat for White people (Goff et al., 2008)
8) Skepticism about social science (adapted from McCright, Dentzman, Charters, & Dietz (2013)
9) Demographics

Study 1 Post-Manipulation (T2)

1. Random assignment to condition
2. Manipulation Checks
    a. Reading comprehension
3. Shortened Motivated Reasoning Battery
    a. Sub-sections:
        i. Belief in Personal Implicit Bias
        ii. Perceptions of Social Scientists and Research
        iii. Belief in the General Pervasiveness and Consequences of Implicit Bias
        iv. Belief in the General Pervasiveness and Consequences of General Prejudice
4. Attitudes towards interventions to reduce implicit bias in organizational contexts, law enforcement, and society generally.
5. MCPR \(Dunton & Fazio, 1997)
6. Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
7. Personal vs. situational attributions about racial disparities (adapted from Peffley & Hurwitz)
8. Attitudes about criminal justice system, affirmative action, and social welfare
9. Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
10. Political ideology/knowledge

Study 2 Pre-Manipulation (T1)

1) Explicit Racial Attitudes
    a. Racial Resentment (Kinder & Sanders, 1996)
    b. Attitudes Toward Blacks Scale (Brigham, 1993)
2) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
3) Authoritarianism (Stenner, 2005)
4) Internal/External Motivation to Control Prejudice (Plant & Devine, 1998)
5) Social Dominance Orientation (Sidanius & Pratto, 2001)
6) System justification (Kay & Jost, 2003)
7) Stereotype threat for White people (Goff et al., 2008)
8) Skepticism about social science (adapted from McCright, Dentzman, Charters, & Dietz (2013)
9) Demographics

Study 2 Post-Manipulation (T2)

1) Random assignment to condition
2) Race IAT
3) Post-IAT Feedback
4) Affect (Devine et al., 1991; Monteith & Voils, 1998; Monteith, Voils, Ashburn-Nardo, 2001)
5) Full Motivated Reasoning Battery
    a. Sub-sections:
        i. Belief in Personal Implicit Bias
        ii. Perceptions of Social Scientists and Research
        iii. Perception/Judgment of IAT Test/Instrument
        iv. Belief in the General Pervasiveness and Consequences of Implicit Bias
        v. Belief in the General Pervasiveness and Consequences of General Prejudice
6) Attitudes towards interventions to reduce implicit bias in organizational contexts, law enforcement, and society generally.
7) MCPR (Dunton & Fazio, 1997)
8) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
9) Personal vs. situational attributions about racial disparities (adapted from Peffley & Hurwitz)
10) Attitudes about criminal justice system, affirmative action, and social welfare
11) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
12) Political ideology/knowledge

Study 3 Pre-Manipulation (T1)

1) Explicit Racial Attitudes
    a. Racial Resentment (Kinder & Sanders, 1996)
    b. Attitudes Toward Blacks Scale (Brigham, 1993)
2) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
3) Authoritarianism (Stenner, 2005)
4) Internal/External Motivation to Control Prejudice (Plant & Devine, 1998)
5) Social Dominance Orientation (Sidanius & Pratto, 2001)
6) System justification (Kay & Jost, 2003)
7) Stereotype threat for White people (Goff et al., 2008)
8) Skepticism about social science (adapted from McCright, Dentzman, Charters, & Dietz (2013)
9) Demographics

Study 3 Post-Manipulation (T2)

1) Random assignment to condition
2) Race IAT
3) Post-IAT Feedback
4) Interventions
5) Affect (Devine et al., 1991; Monteith & Voils, 1998; Monteith, Voils, Ashburn-Nardo, 2001)
6) Full Motivated Reasoning Battery
    a. Sub-sections:
        i. Belief in Personal Implicit Bias
        ii. Perceptions of Social Scientists and Research
        iii. Perception/Judgment of IAT Test/Instrument
        iv. Belief in the General Pervasiveness and Consequences of Implicit Bias
        v. Belief in the General Pervasiveness and Consequences of General Prejudice
7) Attitudes towards interventions to reduce implicit bias in organizational contexts, law enforcement, and society generally.
8) Personal vs. situational attributions about racial disparities (adapted from Peffley & Hurwitz)
9) Attitudes about criminal justice system, affirmative action, and social welfare
10) Instruction test **(**Oppenheimer, Meyvis, Davidenk, 2006)
11) Political ideology/knowledge

Manipulation Checks

The following questions are in regards to your opinion about the article that you just read. Please answer these questions as honestly as you can.

1. The information/feedback presented to me was believable. "Strongly Disagree" (1) to "Strongly Agree" (7)
    a. How confident are you in this? "not at all confident" "extremely confident" (7)
2. I wondered if the information/feedback presented to me was accurate. "Strongly Disagree" (1) to "Strongly Agree" (7)
    a. How confident are you in this? "not at all confident" "extremely confident" (7)
3. What do you think was the true purpose of this experiment? (Open-ended)
4. Do you think the researcher intentionally tried to deceive or mislead you? (Yes/No)
    a. If yes, please explain (open-ended)

Pre-Manipulation Measures

Racial Resentment (Kinder & Sanders, 1996)

1. Over the past few years, blacks have gotten less than they deserve. (Reverse)
2. Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors
3. It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites.
4. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class. (Reverse)

1=agree strongly, 2=agree somewhat, 3=neither agree nor disagree, 4=disagree somewhat, 5=disagree strongly

Attitudes Toward Blacks Scale (Brigham, 1993)

1. If a black were put in charge of me, I would not mind taking advice and direction from him or her.
2. If I had a chance to introduce black visitors to my friends and neighbours, I would be pleased to do so.
3. I would rather not have blacks live in the same neighbourhood I live in. (Reversed)
4. I would probably feel somewhat self-conscious dancing with a black in a public place. (Reversed)

5. I would not mind it at all if a black family with about the same income and education as me moved in next door.
6. I think that black people living in Germany look more similar to each other than white people do. (Reversed)
7. Interracial marriage should be discouraged to avoid the "who-am-I?" confusion which the children feel. (Reversed)
8. I get very upset when I hear a white make a prejudicial remark about blacks.
9. I favour open housing laws that allow more racial integration of neighbourhoods.
10. It would not bother me if my new roommate was black.
11. It is likely that blacks will bring violence to neighbourhoods when they move in. (Reversed)
12. I enjoy a funny racial joke, even if some people might find it offensive. (Reversed)
13. The German governments should take decisive steps to override the injustices blacks suffer at the hands of local authorities.
14. Black and white people are inherently equal.
15. Black people are demanding too much too fast in their push for equal rights in Germany. (Reversed)
16. Whites should support blacks in their struggle against discrimination and segregation in Germany.
17. Generally, blacks are not as smart as whites. (Reversed)
18. I worry that in the next few years I may be denied my application for a job or a promotion because of preferential treatment given to minority group members. (Reversed)
19. Racial integration (of schools, businesses, residences, etc.) has benefitted both whites and blacks.
20. Some blacks in Germany are so touchy about race that it is difficult to get along with them. (Reversed)

Scoring
1 (Strongly Agree)
2 (Agree)
3 (Agree somewhat)
4 (Neutral)
5 (Disagree somewhat)
6 (Disagree)
7 (Strongly Disagree)

Instruction test (Oppenheimer, Meyvis, Davidenk, 2006)

This is just a test to ensure that you are following the instructions. Please click the response option "RED" in the space below.

Authoritarianism (Stenner, 2005)
Please read each pair of qualities that children might have, and indicate which of the two is the most desirable quality for a child to have.

Independence or Respect for Elders:
1. Independence is more desirable
2. Respect for Elders is more desirable
3. I'm not sure
4. Both are equally important
5. Neither is very important

Obedience or Self-reliance:
1. Obedience is more desirable
2. Self-reliance is more desirable
3. I'm not sure
4. Both are equally important
5. Neither is very important

Curiosity or Good Manners:
1. Curiosity is more desirable
2. Good Manners are more desirable
3. I'm not sure
4. Both are equally important
5. Neither is very important

Being Considerate or Well-behaved:
1. Being Considerate is more desirable
2. Being Well-behaved is more desirable
3. I'm not sure
4. Both are equally important
5. Neither is very important

Internal/External Motivation to Control Prejudice (Plant & Devine, 1998)

Internal Motivation to Respond Without Prejudice Scale (IMS) and External Motivation to Respond Without Prejudice Scale (EMS) Items

Instructions: The following questions concern various reasons or motivations people might have for trying to respond in nonprejudiced ways toward Black people. Some of the reasons reflect internal–personal motivations whereas others reflect more external–social motivations. Of course, people may be motivated for both internal and external reasons; we want to emphasize that neither type of motivation is by definition better than the other. In addition, we want to be clear that we are not evaluating you or your individual responses. All your responses will be completely confidential. We are simply trying to get an idea of the types of motivations that students in general have for responding in nonprejudiced ways. If we are to learn anything useful, it is important that you respond to each of the questions openly and honestly. Please give your response according to the scale below.

	Factor loadings	
Scale item	Factor 1: IMS	Factor 2: EMS
External motivation items		
Because of today's PC (politically correct) standards I try to appear nonprejudiced toward Black people.	.05	.73
I try to hide any negative thoughts about Black people in order to avoid negative reactions from others.	−.003	.78
If I acted prejudiced toward Black people, I would be concerned that others would be angry with me.	.22	.67
I attempt to appear nonprejudiced toward Black people in order to avoid disapproval from others.	−.16	.83
I try to act nonprejudiced toward Black people because of pressure from others.	−.22	.69
Internal motivation items		
I attempt to act in nonprejudiced ways toward Black people because it is personally important to me.	.76	.15
According to my personal values, using stereotypes about Black people is OK. (R)	.71	−.16
I am personally motivated by my beliefs to be nonprejudiced toward Black people.	.77	−.08
Because of my personal values, I believe that using stereotypes about Black people is wrong.	.77	−.05
Being nonprejudiced toward Black people is important to my self-concept.	.74	−.08

Note. (R) indicates reverse coded item. Participants rated 10 items on a scale ranging from 1 (*strongly disagree*) to 9 (*strongly agree*). When participants complete the scales, the IMS and EMS items are intermixed. The factor loadings are from an exploratory factor analysis.

Social Dominance Orientation (Sidanius & Pratto, 2001)

INSTRUCTIONS: Using the 7-point scale below, please rate how strongly you agree or disagree with each of the following statements. Circle the number corresponding to your degree of agreement/disagreement with each statement.

Strongly Disagree (1) Disagree (2) Somewhat disagree (3) Neutral or Undecided (4) Somewhat Agree (5) Agree (6) Strongly agree (7)

1 2 3 4 5 6 7

_____ 1. Some groups of people are simply inferior to other groups.

_____ 2. In getting what you want, it is sometimes necessary to use force against other groups.

_____ 3. It is OK if some groups have more of a chance in life than others.

_____ 4. To get ahead in life, it is sometimes necessary to step on other groups.

_____ 5. If certain groups stayed in their place, we would have fewer problems.

_____ 6. It is probably a good thing that certain groups are at the top and other groups are at the bottom.

_____ 7. Inferior groups should stay in their place.

_____ 8. Sometimes other groups must be kept in their place.

_____ 9. It would be good if groups could be equal.

_____ 10. Group equality should be our ideal.

_____ 11. All groups should be given an equal chance in life.

_____ 12. We should do what we can to equalize conditions for different groups.

_____ 13. Increased social equality.

_____ 14. We would have fewer problems if we treated people more equally.

_____ 15. We should strive to make incomes as equal as possible.

_____ 16. No group should dominate in society.

System justification (Kay & Jost, 2003)

1. In general, you find society to be fair.
2. In general, the American political system operates as it should.
3. American society needs to be radically restructured.
4. The United States is the best country in the world to live in.
5. Most policies serve the greater good.
6. Everyone has a fair shot at wealth and happiness.
7. Our society is getting worse every year.
8. Society is set up so that people usually get what they deserve.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

Stereotype threat for White people (Goff et al., 2008)

1. "I worry that my conversation partner might stereotype me as racist because I am White";
2. "I worry that something I say might be misinterpreted as prejudiced by my conversation partner";
3. "I never worry that someone will suspect me of being prejudiced just because I am White"
4. "I worry that my conversation partner's evaluations of me might be affected by my race"
5. "I worry that, because I know the racial stereotype about Whites and prejudice, my anxiety about confirming that stereotype will negatively influence our interaction."

Participants rated their agreement with the statements on a scale from 1 (strongly disagree) to 7 (strongly agree).

Skepticism about social science (adapted from McCright, Dentzman, Charters, & Dietz (2013)

How much confidence would you say you have in the scientific community**?**
1= no confidence at all, 5= complete confidence

'How much do you distrust or trust social scientists to:'

- ‘create knowledge that is unbiased and accurate?’
- ‘create knowledge that is useful?’
- ‘advise government officials on policy?’
- ‘inform the public on important issues?

Scale: 1=completely distrust, 2= partially distrust, 3= neither distrust or trust, 4=partially trust, 5= completely trust

Demographics

Finally, we would like you to give us a little information about yourself. Before completing this questionnaire, please respond to the following background questions.

1. What is your age? _____

2. Your gender? Male Female

3. Please indicate your race/ethnicity.

    a. Latino/Hispanic
    b. Black/African American
    c. Asian/Asian American
    d. White/Caucasian
    e. Native American
    f. Other _____

4. Are you a U.S. citizen? Yes No
5. Were you born in the U.S.A? Yes No
6. What is your total family (including parent income if dependent on parents) income?
    i. _____ Less than $10,000
    ii. _____ $10,000-$19,999
    iii. _____ $20,000-$29,999
    iv. _____ $30,000-$39,999
    v. _____ $40,000-$49,999
    vi. _____ $50,000-$59,999
    vii. _____ $60,000-$69,999
    viii. _____ $70,000-$79,999
    ix. _____ $80,000-$89,999
    x. _____ $90,000-$99,999
    xi. _____ $100,000 or greater

What is the highest level of education you have completed?

1. Grade School
2. Some High School

3. High School Diploma or Equivalent
4. Some College
5. Associate's Degree
6. Bachelor's Degree
7. Master's Degree
8. Advanced Degree (PhD, DPHIL, J.D., M.D., DDS, etc)

Post-Manipulation Measures

Affect (Devine et al., 1991; Monteith & Voils, 1998; Monteith, Voils, Ashburn-Nardo, 2001)

Scale= does not apply at all (1) to applies very much (7).

1) angry at myself
2) guilty
3) irritated with others
4) depressed
5) sad.
6) regretful
7) happy
8) anxious
9) bothered
10) uneasy
11) disgusted with other
12) embarrassed
13) annoyed at myself
14) threatened
15) fearful
16) uncomfortable
17) energetic
18) optimistic
19) content
20) good
21) disappointed with myself
22) disgusted with myself
23) shame
24) self-critical
25) Negative
26) concerned
27) frustrated
28) tense
29) distressed
30) friendly
31) angry at others

Shortened Motivated Reasoning Battery

Please answer the following questions as honestly as you can. Remember, there are no right or wrong answers. We are only interested in your opinion.

Pervasiveness and Consequences of Personal Implicit Bias

1. How likely is it that your unconscious beliefs are unfavorable toward racial minorities? (1=Not at all likely, 7=Extremely Likely)
2. To what extent do you think your unconscious beliefs are unfavorable towards racial minorities? (1=Not at all, 7= Extremely)
3. Do you believe that your unconscious racial attitudes affect your judgments towards racial minorities in an unfair way? (1=Not at all, 7=Yes, definitely)
4. Do you believe that your unconscious racial attitudes influence your behavior towards racial minorities in an unfair way? (1=Not at all, 7=Yes, definitely)
5. Do you believe that you are unconsciously prejudiced towards racial minorities? (1=Not at all, 7=Yes, definitely)
6. How worried are you that you are unconsciously prejudiced towards racial minorities? (1=Not at all worried, 7=Extremely worried)
7. Pretend that your employer wants you and your colleagues to undergo a training program to reduce unconscious racial bias and its consequences. Participation in this program is encouraged, but voluntary.
    a. Would you participate in this anti-bias program? (6-point scale, no mid-point: 1=No, not at all, 6=Yes, definitely)
        i. Please list your reasons for why you would be willing to participate in a program to reduce unconscious racial bias and its consequences. (Open-ended; list up to 5 reasons why or why not)
        ii. Please list your reasons for why you would NOT be willing to participate in a program to reduce unconscious racial bias and its consequences. (Open-ended; list up to 5 reasons why or why not)

Perception of Social Scientists and Research

1. Social scientists think that all White people are racists. (1=Not at all, 7= Yes, definitely)
2. Social scientists think that all Black people are treated unfairly because of their race. (1=Not at all, 7= Yes, definitely)
3. Social scientists think that the only reason Black people don't succeed as much as White people is because they are victims of racial prejudice. (1=Not at all, 7= Yes, definitely)
4. How infected by political motives and values are social scientists? (1=Not at all, 7= Very much so)

5. To what extent are social scientists who study the psychology of unconscious racial bias motivated by a political or ideological agenda? (1=Not at all, 7= Yes, definitely)
6. How credible are social scientists who study the psychology of unconscious racial bias? (1=Not at all credible, 7=extremely credible)
7. How objective are social scientists who study the psychology of unconscious racial bias? (1=Not at all objective, 7=extremely objective)
8. Social scientists specialize in lying with numbers. (1=Strongly Disagree, 7=Strongly Agree)

Pervasiveness and Consequences of General Implicit Bias

1. How common is unconscious racial prejudice in America? (1=Not at all common, 7=extremely common)
2. How likely is it that unconscious racial attitudes biases people's judgments and behavior towards racial minorities?(1=Not at all likely, 7=Extremely Likely)
3. Do you think differences between racial groups can be explained by the effects of unconscious racial bias and prejudice? (1=not at all, 7= definitely)

Pervasiveness and Consequences of General Prejudice (adapted from O'Brien et al., 2010)

1. How common is racial prejudice in America? (1=Not at all common, 7=Extremely common)
2. How big of a problem is racial prejudice in America? (1=Not at all a problem, 7=A major problem)
3. How racially prejudiced is the average American? (1=Not at all prejudiced, 7= extremely prejudiced)
4. Do you think that racial minorities are held back in society because of racial prejudice? ((1=Not at all, 7=Yes, definitely)

Full Motivated Reasoning Battery

Please answer the following questions as honestly as you can. Remember, there are no right or wrong answers. We are only interested in your opinion.

Pervasiveness and Consequences of Personal Implicit Bias

8. How likely is it that your unconscious beliefs are unfavorable toward racial minorities? (1=Not at all likely, 7=Extremely Likely)
9. To what extent do you think your unconscious beliefs are unfavorable towards racial minorities? (1=Not at all, 7= Extremely)

10. Do you believe that your unconscious racial attitudes affect your judgments towards racial minorities in an unfair way? (1=Not at all, 7=Yes, definitely)
11. Do you believe that your unconscious racial attitudes influence your behavior towards racial minorities in an unfair way? (1=Not at all, 7=Yes, definitely)
12. Do you believe that you are unconsciously prejudiced towards racial minorities? (1=Not at all, 7=Yes, definitely)
13. How worried are you that you are unconsciously prejudiced towards racial minorities? (1=Not at all worried, 7=Extremely worried)
14. Pretend that your employer wants you and your colleagues to undergo a training program to reduce unconscious racial bias and its consequences. Participation in this program is encouraged, but voluntary.
    a. Would you participate in this anti-bias program? (6-point scale, no mid-point: 1=No, not at all, 6=Yes, definitely)
        i. Please list your reasons for why you would be willing to participate in a program to reduce unconscious racial bias and its consequences. (Open-ended; list up to 5 reasons why or why not)
        ii. Please list your reasons for why you would NOT be willing to participate in a program to reduce unconscious racial bias and its consequences. (Open-ended; list up to 5 reasons why or why not)

Perceptions/Judgment of IAT Instruments **(STUDY 2 AND 3 ONLY)**

Earlier, you completed a test that involved the pairing of photos of Black and White people with positive/negative traits. Psychologists use this as a test of unconscious racial prejudice.

1. To what extent do you agree with your results from the test? (1=Not at all, 7= extremely; **EXPERIMENTAL CONDITIONS ONLY!!!)**
2. In your opinion, how credible is this test? (1=Not at all credible, 7= Extremely Credible)
3. In your opinion, how objective is this test? (1=Not at all objective, 7= Extremely objective)
4. In your opinion, how valid are the results of this test? (1=Not at all valid, 7= Extremely valid)
5. In your opinion, how useful is this test for understanding people's racial attitudes? (1=Not at all useful, 7= Extremely useful)

Perception of Social Scientists and Research

1. Social scientists think that all White people are racists. (1=Not at all, 7= Yes, definitely)
2. Social scientists think that all Black people are treated unfairly because of their race. (1=Not at all, 7= Yes, definitely)

3. Social scientists think that the only reason Black people don't succeed as much as White people is because they are victims of racial prejudice. (1=Not at all, 7= Yes, definitely)
4. How infected by political motives and values are social scientists? (1=Not at all, 7= Very much so)
5. To what extent are social scientists who study the psychology of unconscious racial bias motivated by a political or ideological agenda? (1=Not at all, 7= Yes, definitely)
6. How credible are social scientists who study the psychology of unconscious racial bias? (1=Not at all credible, 7=extremely credible)
7. How objective are social scientists who study the psychology of unconscious racial bias? (1=Not at all objective, 7=extremely objective)
8. Social scientists specialize in lying with numbers. (1=Strongly Disagree, 7=Strongly Agree)

Pervasiveness and Consequences of General Implicit Bias

4. How common is unconscious racial prejudice in America? (1=Not at all common, 7=extremely common)
5. How likely is it that unconscious racial attitudes biases people's judgments and behavior towards racial minorities?(1=Not at all likely, 7=Extremely Likely)
6. Do you think differences between racial groups can be explained by the effects of unconscious racial bias and prejudice? (1=not at all, 7= definitely)

Pervasiveness and Consequences of General Prejudice (adapted from O'Brien et al., 2010)

5. How common is racial prejudice in America? (1=Not at all common, 7=Extremely common)
6. How big of a problem is racial prejudice in America? (1=Not at all a problem, 7=A major problem)
7. How racially prejudiced is the average American? (1=Not at all prejudiced, 7= extremely prejudiced)
8. Do you think that racial minorities are held back in society because of racial prejudice? ((1=Not at all, 7=Yes, definitely)

Attitudes towards interventions to reduce implicit bias in organizational contexts, law enforcement, and society generally.

1. In your opinion, how important is it that the law requires employers to undergo training to reduce their unconscious racial bias? (1=Not at all important, 7=Extremely important)

2. In your opinion, how important is it that organizations require their employees to undergo training to reduce their unconscious racial bias? (1=Not at all important, 7=Extremely important)
3. In your opinion, how important is it that law enforcement personnel undergo training to reduce their unconscious racial bias? (1=Not at all important, 7=Extremely important)
4. In your opinion, how valuable are anti-bias programs in eliminating racial discrimination in the workplace? (1=Not at all valuable, 7=Extremely valuable)
5. In your opinion, how valuable are anti-bias programs in eliminating racial discrimination in the criminal justice system? (1=Not at all valuable, 7=Extremely valuable)
6. In your opinion, how valuable are anti-bias programs in eliminating racial discrimination in the society? (1=Not at all valuable, 7=Extremely valuable)

Personal vs. situational attributions about racial disparities (adapted from Peffley & Hurwitz)

Statistics show that African-Americans are more often arrested and sentenced to prison than White people. Reasonable people disagree about why this occurs. Some people think that the police and justice system are biased against Blacks, and some people believe that Blacks are more likely to commit crimes and less likely to respect authority than Whites. In your view, why are Blacks more often arrested and sent to prison than Whites?

1	2	3	4	5	6	7
Police and the Criminal Justice System Are Biased Against Blacks						*Blacks are more likely to commit crimes than Whites*

1	2	3	4	5	6	7
Police and the Criminal Justice System Are Biased Against Blacks						*Blacks don't respect authority*

Statistics also indicate that African-Americans are less likely to be hired and promoted in most employment contexts than White people. Reasonable people disagree about why this occurs. Some people think that the employers are biased against Blacks, and some people believe that Blacks are less competent and less hardworking than Whites. In your view, why are Blacks less likely to be hired and promoted than Whites?

1	2	3	4	5	6	7
Employers Are Biased Against Blacks						*Blacks are less competent than Whites*

1	2	3	4	5	6	7
Employers Are Biased Against Blacks						*Blacks are less hardworking than Whites*

Evidence also indicates that African-Americans are more likely to live in poverty than White people. Reasonable people disagree about why this occurs. Some people believe that high rates of poverty in Black communities is due to housing segregation and lack of public assistance. Some people believe that Blacks are too dependent on welfare, and that Black communities do not promote a strong work ethic. In your view, why are Blacks more likely to live in poverty than Whites?

1	2	3	4	5	6	7
Housing segregation and lack of public assistance						*Blacks are too dependent on welfare*

1	2	3	4	5	6	7
Housing segregation and lack of public assistance						*Black communities do not promote a strong work ethic.*

Attitudes about criminal justice system, affirmative action, and social welfare (items taken from ANES; Peffley & Hurwitz, 2010; Gilens, 1997)

Social Welfare

Some people believe that private charities are enough to help the poor, and the government should stay out of the matter. Others believe that citizens will not give enough to private charities to help the poor, and that the government should set up programs to help them. Many also have opinions which fall in between. What is your opinion?

Private charities are enough						Government programs are needed
1	2	3	4	5	6	7

Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Other people feel it is important for the government to provide many more services even if it means an increase in spending. Many also have opinions that fall in between. What is your opinion?

Fewer services; reduce spending						More services; increase spending
1	2	3	4	5	6	7

When people can't support themselves, the government should help by giving them enough money to meet their needs.

Agree strongly	Agree Somewhat	Disagree somewhat	Disagree strongly
1	2	3	4

Criminal Justice

Do you FAVOR or OPPOSE the death penalty for persons convicted of murder?

Favor death penalty						Oppose death penalty
1	2	3	4	5	6	7

Some people say that the best way to reduce crime is to address the social problems that cause crime, like bad schools, poverty and joblessness. Other people say the best way to reduce crime is to make sure that criminals are caught, convicted and punished harshly. Many also have opinions that fall in between. What is your opinion?

Address social problems						Make sure criminals are caught, convicted, and punished harshly
1	2	3	4	5	6	7

In many areas of the country, police officers use a practice known as racial profiling, where they stop and question Black motorists because the officer believe Blacks are more likely to commit certain types of crime. Do you strongly approve, somewhat approve, somewhat disapprove, or strongly disapprove of racial profiling?

Affirmative Action and Employment Discrimination

Some people say that blacks should be given special consideration when applying for jobs and promotions to make up for past discrimination. Other people say that the individual's ability or experience should be the *only* consideration in hiring or promoting people. Many also have opinions which fall in between. How do you feel?

Hiring should be based only on ability						African Americans should be given special consideration
1	2	3	4	5	6	7

Recently, there has been a lot of talk about job discrimination. Do you think that the government should pass laws to protect racial minorities from job discrimination?

Government should *not* protect homosexuals from job discrimination						Government should protect homosexuals from job discrimination
1	2	3	4	5	6	7

Should the government in Washington see to it that Black people get fair treatment in jobs or is this not the federal government's business

Government should see to fair treatment						Not the government's business
1	2	3	4	5	6	7

MCPR (Dunton & Fazio, 1997)

Instructions: Please indicate your level of agreement with each of the following statements by circling your response on the scale below.

In today's society it is important that one not be perceived as prejudiced in any manner.

1----------2----------3----------4----------5----------6----------7
Strongly Agree — Neither Agree Nor Disagree — Strongly Disagree

I always express my thoughts and feelings, regardless of how controversial they might be. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I get angry with myself when I have a thought or feeling that might be considered prejudiced.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

If I were participating in a class discussions and a Black student expressed and opinion with which I disagreed, I would be hesitant to express my own viewpoint.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

Going through life worrying about whether you might offend someone is just more trouble than it's worth. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

It's important to be that other people not think I'm prejudiced.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I feel it's important to behave according to society's standards.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I'm careful not to offend my friends, but I don't worry about offending people I don't know or like. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I think that it is important to speak one's mind rather than to worry about offending someone (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

It's never acceptable to express one's prejudices.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I feel guilty when I have a negative though or feeling about a Black person.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

When speaking to a Black person, it's important to me that he/she not think I'm prejudiced.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

Instruction test (Oppenheimer, Meyvis, Davidenk, 2006)

This is just a test to ensure that you are following the instructions. Please click the response option "6" in the space below.

Political Partisanship/Ideology

We'd like you to respond to the following questions about your political beliefs.

1. How would you describe your **political party preference** (circle one)?

1 Strong Democrat
2 Weak Democrat
3 Independent/Lean Democrat
4 Independent
5 Independent/Lean Republican
6 Weak Republican
7 Strong Republican

2. To what extent do you feel certain about your political party preference?

1	2	3	4	5
Not at all				Very much

3. How would you describe your ideological preference **in general** (circle one)?

1	2	3	4	5	6	7
Very liberal	Liberal	Slightly Liberal	Moderate	Slightly Conservative	Conservative	Very Conservative

4. To what extent do you feel certain about your liberal/conservative outlook?

1	2	3	4	5
Not at all				Very much

Political Knowledge

We would like to ask you a few questions about public figures and the political system in general. Please respond to each of the following questions as thoroughly as possible.

1. What job or political office does **Joseph Biden** currently hold?
    a. Attorney General
    b. Vice President
    c. Secretary of State
    d. Speaker of the House
    e. Governor of New Hampshire

2. What job or political office does **John Roberts** currently hold?
    a. Secretary of Defense
    b. Attorney General
    c. Senate Majority Leader
    d. Secretary of the Interior
    e. Chief Justice of the Supreme Court

3. What job or political office does **David Cameron** currently hold?
    a. Speaker of the United Nations' General Assembly
    b. Prime Minister of the United Kingdom
    c. Prime Minister of Australia
    d. U.S. envoy to the United Nations
    e. Head of the European Commission

4. What job or political office does **John Boehner** currently hold?
    a. Speaker of the House
    b. Secretary Treasury
    c. Secretary of Homeland Security
    d. White House Chief of Staff
    e. Attorney General

5. Which political party currently has the most members in the Senate in Washington?
    a. Democrats
    b. Republicans
    c. Both parties have the same number of members

6. Which political party currently has the most members in the House of Representatives in Washington?
    d. Democrats
    e. Republicans
    f. Both parties have the same number of members

7. How long is the term of office for a U.S. senator?
    a. 2 years
    b. 4 years
    c. 5 years
    d. 6 years
    e. 8 years

8. Whose responsibility is it to nominate judges to the Federal Courts — the President, the Congress, or the Supreme Court?
    a. The President
    b. Congress
    c. The Supreme Court

Pilot/ Study 1 Post-Manipulation Questions

Reading Comprehension Questions for Experimental Condition:

1. (T/F) Social and behavioral scientists conclude that most Americans are unconsciously prejudiced towards racial minorities.
2. (T/F) A large body of scientific evidence proves that even Americans who intend well and support racial inequality are unconsciously prejudiced towards racial minorities.
3. (T/F) Social and behavioral scientist believe that unconscious racial prejudice is the primary cause of racial inequality in our society.
4. Which of the following statements is most consistent with the consensus opinion of social and behavioral scientists?
    a. All Americans are unconsciously racist.
    b. Most Americans are unconsciously racist.
    c. Some Americans are unconsciously racist.
    d. Few Americans are unconsciously racist.
    e. None of the above.

Reading Comprehension Questions for Control Condition:

1. (T/F) Social and behavioral scientists conclude that most Americans who exercise and eat healthy are happy.
2. (T/F) A large body of scientific evidence proves that Americans who eat healthy and exercise regularly are less stressed, more energetic, and better able to cope with illness.
3. (T/F) Most social and behavioral scientists believe that though most American want to be happy, few adopt the necessary changes to achieve that goal.

4. Which of the following statements is most consistent with the consensus opinion of social and behavioral scientists?
    a. All Americans who exercise and eat healthy are happy.
    b. Most Americans who exercise and eat healthy are happy.
    c. Some Americans who exercise and eat healthy are happy.
    d. Few Americans who exercise and eat healthy are happy.

Study 3 Interventions

Self-Affirmation Intervention

Instructions: Please indicate your level of agreement with each of the following statements by circling your response on the scale below.

In today's society it is important that one not be perceived as prejudiced in any manner.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I always express my thoughts and feelings, regardless of how controversial they might be. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I get angry with myself when I have a thought or feeling that might be considered prejudiced.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

If I were participating in a class discussions and a Black student expressed and opinion with which I disagreed, I would be hesitant to express my own viewpoint.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

Going through life worrying about whether you might offend someone is just more trouble than it's worth. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

It's important to be that other people not think I'm prejudiced.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I feel it's important to behave according to society's standards.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I'm careful not to offend my friends, but I don't worry about offending people I don't know or like. (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I think that it is important to speak one's mind rather than to worry about offending someone (R)

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

It's never acceptable to express one's prejudices.

1----------2----------3----------4----------5----------6----------7
Strongly Agree | Neither Agree Nor Disagree | Strongly Disagree

I feel guilty when I have a negative thought or feeling about a Black person.

1----------2----------3----------4----------5----------6----------7

Strongly Agree	Neither Agree Nor Disagree	Strongly Disagree

When speaking to a Black person, it's important to me that he/she not think I'm prejudiced.

1----------2----------3----------4----------5----------6----------7

Strongly Agree	Neither Agree Nor Disagree	Strongly Disagree

Ranking of Personality Characteristics and Values

Instructions: Below is a list of characteristics and values, some of which may be important to you and some of which may be unimportant. Please rank these values and qualities in order of their importance to you, from 1 to 12.

1 = most important item, 12 = least important item. Use each number only once.

My values and qualities:

_____ Treating people fairly

_____ Artistic skills/Aesthetic appreciation

_____ Sense of humor

_____ Relations with friends/family

_____ Spontaneity/living life in the moment

_____ Social skills

_____ Athletics

_____ Music ability/appreciation

_____ Neatness/tidiness

_____ Physical attractiveness

_____ Creativity

_____ Business/managerial skills

_____ Being kind to others

On this page, please indicate what value you ranked # 1 in the previous exercise. Then, write a brief account (1-3 paragraphs) of why this value is important to you and a time when your 1st-ranked value played an important role in your life.

Number 1 value: _____

(new page)

It is important to understand that this test does NOT *guarantee* that you are racially biased, nor does it mean that you have discriminated against racial minorities in the past. While unconscious bias is extremely common and is quite normal, it is also something that, once you are aware of it, you are able to control. For example, this test has been administered to a very large sample of the people in several different studies. The results from these studies indicate that the overwhelming majority of people harbor unconscious racial bias- even among people who strongly support racial equality and value racial tolerance. However, people who were made aware of their implicit bias were also better able to control it and minimize its influence on their judgment and behavior.

Social and behavioral scientists agree that unconscious preferences for some racial groups are a normal, basic feature of human cognition, and have been reliably observed across most cultures and historical periods. In fact, one study determined that even social scientists who study racial discrimination commonly harbor unconscious racial prejudice. Most psychologists believe that unconscious beliefs, like the beliefs measured by this test, reflect the information available in the social environment and not some deep-rooted bigotry or hatred towards people in society. In this sense, unconscious racial bias is a basic feature of human cognition, It is a common and normal consequence of living in modern times, but it also something that people are able to control, once they become aware that it is influencing their thoughts and behavior.

1. (T/F) According to psychological scientists, unconscious prejudice is extremely common in the American population.
2. (T/F) Most psychological scientists agree that unconscious racial prejudice in basic feature of human cognition.
3. (T/F) Prior research indicates that even social scientists who study race relations harbor unconscious racial prejudice.
4. (T/F) Unconscious beliefs reflect the information in the social environment, and not some deep-rooted bigotry or hatred towards racial minorities.

Open-Ended Coding Scheme for Pilot 2

Reaction to IAT Test (different from reactions to information and feedback)

General Criteria: Participant mentions the "test", "sorting task", "measure of race", "reactions" or something that could be understood as a reference to the Implicit Association Test (IAT), an unconscious measure of racial attitudes.

Exclusion Criteria: If the participant mentions race, stereotypes, or intergroup attitudes, other categories may apply instead. Avoid redundancy between this category and categories below.

Examples Include:

-Skeptical of what the test is measuring
-doubt it's credibility or validity
-criticize the IAT?
-" to see if the brain held on to previous links such as white good, black bad and could adjust to new parameters"
-"Determining how quickly someone can switch tracks in their brains"
-"To see how fast our mind can handle a change in associations."
-"e key is on the left the I key is on the right. and I know how to focus."

Deceive or Manipulation Beliefs

General Criteria: Participant mentions the word "face", "fabricate", "deceive", "manipulate", "brainwashing" or some other statement related to the use of deception or skepticism of the accuracy of the information or feedback they encountered.

Examples Include:

-"fake, fabricated, false information (in general vs. about race)"
-"results not real"
-"Information not real"
-"To make me believe I did better than I think I did."
-"can't be right"
-"to see how easily someone can be deceived."
-"Brainwashing"

To Study Prejudice and Racial Attitudes

General Criteria: Participant mentions the word "race", "prejudice", "bias", "discrimination", "privilege" or some other concept related to racial prejudice and/or inequality.

Exclusion Criteria: If the participant mentions unconscious mode of psychological functioning, the category above may apply instead. If the participant discusses information, feedback, or science, the category below may also apply. Err on the side of applying this category when content is relevant. *It's okay to apply both this category and categories above/below.*

Note: In addition to applying the overall category "Study Prejudice and Racial Attitudes", the following subcategories should also be applied, when appropriate. However, you can apply multiple subcategories or no subcategories, even if the content belongs to the general category.

Subcategories

1) Learn about people biases, prejudice and racial attitudes.
2) Study reactions to general information about race.
3) Study reactions to feedback about one's own racial attitudes
    a. Including information that one is or is not racially biased.

To Deceive About Prejudice and Racial Attitudes

General Criteria: Participant mentions the word "race", "prejudice", "bias", "discrimination", "privilege" or some other concept related to racial prejudice and/or inequality.

Exclusion Criteria: If the participant mentions unconscious mode of psychological functioning, the category above may apply instead. If the participant discusses information, feedback, or science, the category below may also apply. Err on the side of applying this category when content is relevant. *It's okay to apply both this category and categories above/below.*

Note: In addition to applying the overall category "Study Prejudice and Racial Attitudes", the following subcategories should also be applied, when appropriate. However, you can apply multiple subcategories or no subcategories, even if the content belongs to the general category.

Subcategories

1) To see if people believe in racism.
2) Influence opinions about race.
3) Study perception of race.
4) Get people to recognize and admit personal prejudice.
5) Convince people that white's are racists
    a. This includes reference to "white privilege"