Georgia State University
ScholarWorks @ Georgia State University

Psychology Dissertations | Department of Psychology

8-18-2010

Reducing Automatic Stereotype Activation: Mechanisms and Moderators of Situational Attribution Training

Ioana M. Latu
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/psych_diss

Recommended Citation
Latu, Ioana M., "Reducing Automatic Stereotype Activation: Mechanisms and Moderators of Situational Attribution Training" (2010). *Psychology Dissertations.* Paper 72.

This Dissertation is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

REDUCING AUTOMATIC STEREOTYPE ACTIVATION: MECHANISMS AND MODERATORS OF SITUATIONAL ATTRIBUTION TRAINING

by

IOANA MARIA LATU

Under the Direction of Tracie L. Stewart

ABSTRACT

Individuals tend to underestimate situational causes and overly rely on trait causes in explaining negative behaviors of outgroup members, a tendency named the ultimate attribution error (Pettigrew, 1979). This attributional pattern is directly related to stereotyping, because attributing negative behaviors to internal, stable causes tends to perpetuate negative stereotypes of outgroup members. Recent research on implicit bias reduction revealed that circumventing individuals' tendency to engage in the ultimate attribution error led to reduced stereotyping. More specifically, training White participants to consider situational factors in determining Blacks' negative stereotypic behaviors led to decreased automatic stereotype activation. This technique was named Situational Attribution Training (Stewart, Latu, Kawakami, & Myers,

2010). In the current studies, I investigated the mechanisms and moderators of Situational Attribution Training. In Study 1, I investigated the effect of training on spontaneous situational inferences. Findings revealed that training did not increase spontaneous situational inferences: both training and control participants showed evidence of spontaneous situational inferences. In Study 2, I investigated whether correcting trait inferences by taking into account situational factors has become automatic after training. In addition, explicit prejudice, motivations to control prejudice, and cognitive complexity variables (need for cognition, personal need for structure) were investigated as moderators of training success. These findings revealed that Situational Attribution Training works best for individuals high in need for cognition, under conditions of no cognitive load, but not high cognitive load. Training increased implicit bias for individuals high in modern racism, regardless of their cognitive load. Possible explanations of these findings were discussed, including methodological limitations and theoretical implications.

INDEX WORDS: Stereotyping, Stereotype reduction, Automaticity, Individual differences

REDUCING AUTOMATIC STEREOTYPE ACTIVATION: MECHANISMS AND MODERATORS OF SITUATIONAL ATTRIBUTION TRAINING

by

IOANA MARIA LATU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2010

Copyright by

Ioana Maria Latu

2010

REDUCING AUTOMATIC STEREOTYPE ACTIVATION: MECHANISMS AND MODERATORS OF SITUATIONAL ATTRIBUTION TRAINING

by

IOANA MARIA LATU

Committee Chair: Tracie L. Stewart

Committee: Heather Kleider
Dominic Parrott
Ann Pearman
David Washburn

Electronic Version Approved: 07/08/2010

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2010

I would like to dedicate this dissertation to my husband, Michael Triplett, who has been with me for seven years, which coincide with the seven years I have been in graduate school. He believed in me and supported me in every way possible through these years, including taking care of our newborn twin daughters while I was finishing this dissertation. I am sure he's looking forward to my graduation.

ACKNOWLEDGEMENTS

I would like to thank my mentor Tracie Stewart who has been an amazing advisor, teacher, and role model. I wouldn't have gotten here without her support. I would also like to thank my committee members Heather Kleider, Ann Pearman, Dominic Parrott, and David Washburn for offering their expertise in designing and conducting this research. Thank you to Triman Kohli, Glenna Read, Robert Thomas, Erin McFry, and Kerstin Herrmann-Olczak for assistance with data collection.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **v**

LIST OF TABLES ..... **viii**

CHAPTER 1: REVIEW **1**

Implicit Bias **2**

The Roots of Implicit Bias Research **2**

Theoretical Models of Implicit Bias **5**

Measures of Implicit Bias **8**

Implicit Bias and Behavior **14**

Implicit Bias Reduction **17**

Automaticity and Social Inferences **26**

Automatic and Controlled Processes in Social Cognition **26**

Assessing Automaticity in Social Cognition **29**

Spontaneous Inferences **33**

Individual Differences and Implicit Bias **35**

Explicit Prejudice **35**

Motivations to Control Prejudice **36**

Cognitive Complexity **37**

The Present Research **39**

Study 1 ..... **40**

Study 2 ..... **42**

Hypotheses Overview **47**

CHAPTER 2: METHOD **50**

Study 1 ..... **50**

Participants and Design **50**

Procedure and Materials **51**

Results **57**

Discussion **58**

Study 2 ..... **62**

Participants and Design **62**

Procedure and Materials **63**

Results **69**

Discussion **89**

CHAPTER 3: GENERAL DISCUSSION **98**

Hypotheses Review **98**

Overall Implications of Findings **99**

Future Directions **105**

Training Applications **108**

Conclusion *109*

ENDNOTES **112**

REFERENCES **114**

APPENDIX A - CONDITIONS **130**

APPENDIX B - SCALES **131**

LIST OF TABLES

Table 1. Cognitive Load Manipulations in the Literature 31

Table 2. Behaviors Used in the Probe Recognition Task (Study 1), Associated Situations and Traits, and Percentage of Pretest Participants Who Generated Each Situation (N = 22) and Trait (N = 18) 53

Table 3. Example of a behavior and associated probes used in Study 1 56

Table 4. Behaviors Used in the Probe Recognition Task (Study 2), Associated Traits, and Percentage of Pretest Participants Who Generated Each Trait (N = 18)..... 67

Table 5. Example of a behavior and associated probes used in Study 2 69

Table 6. Means, Standard Deviations, and Intercorrelations for Situational Attribution Training Participants who completed the probe recognition task under No Cognitive Load (Above the Diagonal) and Cognitive Load (Below the Diagonal) Conditions in Study 2, N = 47 73

Table 7. Means, Standard Deviations, and Intercorrelations for Grammar Control Participants who completed the probe recognition task under No Cognitive Load (Above the Diagonal) and Cognitive Load (Below the Diagonal) Conditions in Study 2, N = 41 74

Table 8. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Need for Cognition (NFC), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 76

Table 9. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Need for Cognition (NFC), Training Condition, and Load Condition Predicting Stereotyping Scores, the Difference Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 78

Table 10. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Modern Racism (MRS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 80

Table 11. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Modern Racism (MRS), Training Condition, and Load Condition Predicting the Difference Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 82

Table 12. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Internal Motivation to Control Prejudice (IMS), External Motivation to Control Prejudice (EMS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 84

Table 13. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Internal Motivation to Control Prejudice (IMS), External Motivation to Control Prejudice (EMS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)..... 87

Figure 1. Stereotyping Level by Level of Need for Cognition (NFC), Training and Load Condition..... 96

Figure 2. Stereotyping Level by Training Condition and Initial Level of Modern Racism..... 97

Figure 3. The Mechanism Behind Situational Attribution Training 111

CHAPTER 1: REVIEW

In the current research I investigated the effectiveness of a training technique designed to reduce implicit racial bias. The Situational Attribution Training (Stewart, Latu, Kawakami, & Myers, 2010) technique is an implicit bias reduction method rooted in the theory behind the ultimate attribution error (Pettigrew, 1979) – the tendency to attribute negative stereotypic behaviors of outgroup members to dispositional causes, while underestimating the impact of situational causes. In two studies, Stewart and colleagues found that teaching White participants to consider situational causes of Black men's negative behaviors led to reduced implicit bias. In the current research, I focused on three main research questions related to the effectiveness of Situational Attribution Training: whether training has an effect on situational inferences, whether the effects of training are automatic, and whether individual differences moderate the effects of training on implicit bias reduction.

It is important to consider the mechanisms and moderators of Situational Attribution Training in the context of the larger literature on implicit bias, automaticity in social cognition, and individual differences. Thus, the current chapter discusses the theoretical grounds on which research on Situational Attribution Training was built. In the first part I discuss issues surrounding the concept of implicit bias, such as the roots of implicit bias research, theoretical models and methodological paradigms used to measure implicit bias, the relationship between implicit bias and discriminatory behavior, as well as the effectiveness of implicit bias reduction techniques. In the second part of this chapter I discuss the issue of automaticity in social cognition, including research on assessing automaticity through cognitive load tasks as well as research on spontaneous inferences. In the third part of the chapter I define and present research on individual differences that may moderate the effects of Situational Attribution Training on

implicit bias reduction. This part includes a discussion of explicit prejudice, motivations to control prejudice, and cognitive complexity variables such as need for cognition (NFC) and personal need for structure (PNS). Finally, the present chapter ends with an overview of the research questions, methods, and specific hypotheses of the current research.

Implicit Bias

For more than two decades, social psychologists focusing on stereotyping and prejudice have been interested in the study of implicit bias – associating certain social groups (e.g., women, African Americans) with certain traits (e.g., incompetent, lazy) or words that are negative in valence (e.g., filth, vomit). Despite some controversies (Fazio & Olson, 2003; Gawronski, LeBel, & Peters, 2007), these implicit attitudes and stereotypes are generally seen as automatic: they occur without the person's awareness, intention, and control, and they are highly efficient, in the sense that they require few attentional resources to be activated (Bargh, 1994).

The Roots of Implicit Bias Research

As suggested by Amodio and Menodza (in press), research on implicit attitudes in social psychology has two main roots: practical concerns about self-report (explicit) attitudes not being able to predict behavior, as well as theoretical and methodological advances in other fields, particularly in cognitive psychology.

First, early in the stereotype and prejudice research, there seemed to be a disconnect between attitudes and behavior. For example, the famous Princeton Trilogy, a series of three studies conducted to assess the content of African American stereotypes for college students,

suggested that negative racial stereotypes were on the decline. In 1933, 75% of Katz and Braly's sample explicitly stated that African Americans are lazy, followed by 31% in 1951 in Gilbert's sample, and 26% in 1969 in Karlins, Coffman, and Walter's sample. In 2001 (Madon et al.) these percentages decreased even more: 12.1 and 0% in two samples of European American students. The same linear decrease was visible for other negative stereotypic traits such as superstitious, ignorant, ostentatious, and stupid. Despite these encouraging findings, instances of discriminatory behavior were still documented in research. For example, Kempf and Austin (1986) found consistent evidence of racial discrimination in sentencing, based on an analysis of 2,907 tried cases in Pennsylvania in 1977. Juni, Brannon, and Roth (1988) reported a same-race bias in consumer interactions, with White customers preferring White to Black cashiers. In a more direct test of the relationship between racial attitudes and behavior, Rankin and Campbell (1955) revealed that White college students had more negative physiological reactions to a Black compared to a White experimenter, as measured by galvanic skin-responses, despite reporting similarly positive attitudes for Blacks and Whites. Given this discrepancy between attitudes and behaviors, researchers began having a general skepticism about self-report measures and proposed two possible explanations. First, it may be that participants do not want to show their true attitudes about different social groups, due to social desirability concerns (Paulhus, 1984). Alternatively, they may not be aware of their true attitudes, because their introspective skills are limited and often guided by inaccurate naïve theories of their own behavior. The use of implicit measures sought to address both of these issues.

A second source of inspiration for the study of implicit attitudes and stereotypes came from advances in the field of cognitive psychology, particularly learning and memory.

Theoretically, cognitive psychologists advanced the idea that category processing may be automatic, such that semantically related items are judged faster than non-related items. Methodologically, new techniques were developed, such as the sequential priming technique, which allowed for the measurement of semantic associations through reaction time tasks, as opposed to self-report or explicit measures (Meyer & Schvaneveldt, 1971, 1976). For example, in one early study, participants made faster same-different judgments of letter strings when the letter strings formed words that were semantically related compared to when they formed non-words or words that were not semantically related (Meyer & Schvaneveldt, 1971). This finding is in line with predictions from spreading activation models: the activation of the prime spreads to semantically related concepts, thus reducing the time required to activate them. These semantic priming tasks inspired social psychologists to measure social attitudes without using self-report, by measuring semantic associations between certain words (positive versus negative, stereotypic versus non-stereotypic) and social groups. Although many different tasks were later developed in the field of implicit bias, the underlying principle is the same: the more two concepts are associated in a person's mind, the faster the person is at responding to a variety of quick decisions which pair those two concepts together.

It is important to note that there is some controversy in the literature about the terms used to describe these implicit concepts and their measures. For example, Greenwald and Banaji (1995) are very comprehensive in their definition of these terms and equate the explicit-implicit dichotomy with terms such as conscious-unconscious, aware-unaware, direct-indirect, and controlled-automatic. Other researchers contend that there is not enough evidence to call the implicit measures unconscious, unaware, or automatic, and suggest that most of these measures are just indirect measures (Fazio & Olson, 2003; Gawronski, LeBel, & Peters, 2007) – meaning

that participants are usually unaware that their attitudes are being measured, but not necessarily unaware of their attitudes. In the remainder of this manuscript, I will use the terms “implicit,” as it is widely used and accepted in the literature.

Theoretical Models of Implicit Bias

Despite Fazio and Olsen’s (2003) claim that the implicit bias literature is largely non-theoretical, but methodologically and empirically driven, there are several theoretical models in the literature that explain the relationship between implicit and explicit attitudes. In most part, these are dual-processing models, which address the issues of automatic versus controlled processing in forming attitudes about social groups.

One of the first dual-processing models and empirical studies in the field of stereotyping and prejudice comes from Patricia Devine in 1989. Devine makes a distinction between knowledge of stereotypes and endorsement of stereotypes. Knowledge of stereotypes is associated with automatic processing and has its source in the unintentional activation of well-learned associations derived from cultural stereotypes. Endorsement of stereotypes is associated with controlled processing – a person can either consciously endorse or reject those cultural stereotypic associations. As such, both low- and high-prejudice participants would show similar levels of implicit bias. For both groups, whether they endorse it or not, the cultural stereotype of Blacks would be activated when encountering a Black person. Thus, the difference between high- and low-prejudice individuals lies in the overlap between their implicit and explicit associations and beliefs. For low-prejudice individuals, there is little overlap between implicit and explicit attitudes, as these individuals consciously reject the negative associations derived from the cultural stereotype. High-prejudice individuals, on the other hand, show a greater

overlap, as they consciously endorse the cultural stereotype. These hypotheses were supported by several of Devine's studies. For example, in one study, both high- and low-prejudice participants rated a man's ambiguous behavior as more hostile after being primed with the Black stereotype of hostility compared to a condition in which they were not primed with this stereotype. This finding suggests that stereotypes are automatically activated and used in impression formation, regardless of participants' level of explicit prejudice.

Devine's model was later challenged by Fazio's MODE model of attitude behavior processes (Fazio & Towels-Schwen, 1999). Also a dual-processing model, this theoretical framework differentiates between controlled, theory-driven processing and spontaneous, data-driven processing. Fazio and his colleague propose that whether a person engages in automatic or controlled processing depends on two factors: motivation and opportunity (hence the name of the model: Motivation and Opportunity as DEterminants). Motivation refers to a personal motivation to be accurate and to reduce negative consequences when processing social information. Opportunity refers to the time and resources to deliberate. The greater the motivation and opportunity, the more likely an individual is to engage in controlled, deliberate processing. When applied to racial prejudice, the MODE model predicts that we not only differ in the extent to which we explicitly endorse stereotypes, as Devine (1989) suggested, but we also differ in the degree to which we have implicit stereotypes. Fazio and Towels-Schwen regard these automatic structures as personal attitudes, and not shared attitudes and stereotypes, as suggested by Devine. In a series of studies, Fazio and his colleagues investigated the hypothesis that motivation and opportunity moderate the relationship between implicit and explicit racial attitudes. Their finding show that as motivation to control prejudice decreased, the positive relationship between explicit and implicit bias grew stronger. For those who had

positive implicit racial attitudes, motivation mattered very little. However, for those with negative automatic attitudes, motivation played a large role in determining their explicit attitudes, such that they had impressively positive explicit attitudes towards Blacks. Thus, contrary to Devine (1989), Fazio and his colleagues found that individuals not only differ in their explicit, but also implicit attitudes.

More recently, Gawronski and Bodenhausen (2006) proposed a comprehensive model that combines several characteristics of past models. According to their APE model, there are two types of processing styles – Associative and Propositional Evaluations. Associative evaluations are automatic affective reactions that do not require cognitive capacity or intention. They are independent of the assignment of truth values and are not necessarily personally endorsed. Propositional evaluations are cognitive judgments, which are superordinate to the associative store, meaning that this propositional system can take associations and transform them into propositional processes, subjecting them to validity checks. Gawronski and Bodenhausen thus propose that associative judgments can be a source of evaluative judgments, but, more importantly, that this relationship is bidirectional, such that propositions can be a source of associations too. The mere knowledge of a proposition endorsed by other people, even if it is deemed invalid by the person, can contribute to the activation of corresponding associations in memory. This view is more consistent with Devine's (1989) postulation that negative stereotypes can be activated regardless of a person's conscious endorsement of these associations. Regarding the conscious-unconscious debate in explicit-implicit attitudes, the APE model suggests that although explicit attitudes tend to be conscious, and implicit attitudes unconscious, implicit affective associations can sometimes be in a person's awareness. Finally, Gawronski and Bodenhausen claim that both implicit and explicit attitudes are online

constructions, such that they are not stable representations, but instead formed and modified online depending on the context.

The theory behind dual-process models is not the only focus of the implicit bias literature. Also of interest is the effort to differentiate between various types of implicit biases and their measures. For example, Amodio and Devine (2006) suggest that commonly used implicit measures usually assess two independent constructs: implicit stereotyping and implicit evaluations (attitudes), and that this distinction has important consequences for these measures' ability to predict behavior. Implicit stereotyping is a cognitive construct, based on semantic learning and memory. It entails associating different social groups (e.g., African Americans) with stereotypic traits (e.g., lazy, uneducated, athletic, religious). These semantic associations are independent of affective or evaluative associations (positive or negative). Implicit evaluation is an affective construct, based on reward and punishment cues. It entails associating social groups (e.g., African Americans) with positive or negative evaluative words (e.g., cancer, death), which are usually independent of semantic associations. Social neuroscience findings support the independence of these two constructs (see Amodio, 2008, for a review). Implicit stereotyping is usually localized in neocortical structures such as the prefrontal cortex, an area usually associated with semantic processing. Implicit evaluations are localized in the amygdala, an area usually associated with affective processes.

Measures of Implicit Bias

Sequentially primed decision tasks. Many implicit bias measures are computer-based tasks that measure participants' response time to a sequentially primed decision task. For example, in the person categorization task (Blair & Banaji, 1996), participants are presented

briefly with a trait (either positive or negative and either Black stereotypic or non-stereotypic), followed by a photo of either a Black or a White man. Their task is to decide as quickly as possible whether the person in the photo is Black or White, by pressing designated keys on the keyboard. Their response time to this decision is recorded in milliseconds. Participants exhibit automatic activation of negative Black stereotypes if they are faster at correctly categorizing Black compared to White photos after being primed with negative Black stereotypic traits. In a similar paradigm, the lexical decision task (Wittenbrink, Judd, & Park, 1997), the trait and photo presentation is reversed, such that participants have to decide quickly whether certain traits are words or non-words after being primed with Black and White-related stimuli, such as words, symbols, or pictures. Faster response times to categorizing negative Black-stereotypic traits after being primed with Black versus White photos suggests automatic activation of negative Black stereotypes. For both of these tasks, the underlying logic is similar: faster response times to categorization decisions suggest that the stereotypical trait and the target are part of a common semantic network, which facilitates rapid decision making.

The opposite of facilitation, namely inhibition, is the underlying principle of a different task, the modified Stroop task (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). Participants are presented with a category prime (e.g., “elderly” and “skinhead”) followed by elderly and skinhead stereotypes, presented in different colors on the screen. Participants’ task is to state the color of the word into a microphone. In the Kawakami and colleagues study, several dependent measures were recorded, such as response time to naming the color and errors in naming the color. The latter included stutters, mispronunciations, stating the wrong color, or changes in voice characteristics. To the extent that the elderly stereotype was activated in response to the category prime, participants should take longer and/or make more errors in

naming the color of the elderly-stereotypic word, due to inhibition coming from the activation of the elderly stereotype.

By far, the most common implicit measure employed by social psychologists is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). Similar to other tests, the IAT assesses the degree of association between social groups (e.g., male/female) and certain concepts (e.g., good/bad), by measuring participants' reaction times to categorizing items of each category. For example, the gender attitude IAT uses trials in which participants categorize words (e.g., he, she, flower, filth) into pairs of categories displayed on each side of the screen, such as "Male or Good" on the left side and "Female or Bad" on the right side. The more that participants associate men with positive characteristics and women with negative characteristics, the faster they are at categorizing the items, compared to other trials in which the category pairings are switched on the computer screen.

A more specific task is the weapons identification task (Payne, 2001), designed to measure the degree of association between racial groups and weapons. Participants categorize pictures of weapons or hand tools after being primed with photos of Black and White individuals. Both response time and accuracy measures showed racial bias in the perception of weapons, such that participants who were primed with Black versus White photos were faster to identify a gun and were more likely to misidentify a tool as a gun.

Correll and his colleagues (Correll, Park, Judd, & Wittenbrink, 2002) proposed a similar, although more ecologically valid method of measuring racial bias in relation to decisions to shoot. In their shooter task paradigm, participants are presented with pictures of Black and White men carrying either guns or objects other than guns (cell phones, wallets).

Their task is to press a button labeled “Shoot” whenever the target is carrying a gun and a button labeled “Don’t shoot” whenever the target is carrying an object other than a gun. In several studies, participants showed a racial shooter bias both in terms of hits (correctly shooting an armed target) and false alarms (incorrectly shooting an unarmed target). In short, participants were faster to shoot Black compared to White armed targets and more likely to inaccurately shoot Black compared to White unarmed targets.

Memory tasks. Given that implicit bias means automatically associating social groups with certain negative traits, one way to measure bias would be to measure spontaneous trait inferences (STIs) - how people infer the causes of other’s behaviors, by using certain traits. Several studies (Lupfer, Clark, Church, DePaola, McDonald, 1995; Stewart, Weeks, & Lupfer, 2003; Uleman, Hon, Roman, & Moskowitz, 1996a) showed that people tend to spontaneously make trait inferences about others’ behaviors, especially when the behavior is consistent with the stereotype of the group that the person belongs to (Wigboldus, Sherman, Franzese, & van Knippenberg, 2004). One way to study STIs is the probe recognition task (McKoon & Ratcliff, 1986). In this task, participants are presented with several behavioral sentences, such as “Larry lost his job.” After each sentence a probe word is presented on the screen. Participants’ task is to decide whether the word appeared in the sentence or not. On the experimental trials, the probe is a trait that was implied by the behavioral sentence, such as “incompetent.” This probe did not actually appear in the sentence, so the correct answer is “NO.” Participants also respond to a control probe, which did not appear and was not implied by the sentence. To the extent that the trait inference was spontaneously activated when reading the behavioral sentence, participants should have a harder time responding correctly to the trait compared to the control probe. Previous studies using this task found that people tend to spontaneously make trait

inferences, as suggested by both accuracy data – being more likely to identify incorrectly that a trait appeared in the sentence compared to a control (Lupfer et al., 1995; Stewart et al, 2003; Uleman et al, 1996a) and reaction time data– being slower to reject a trait compared to a control (Uleman et al, 1996a, Wigboldus et al., 2004).

Research also looked at whether people tend to spontaneously make situational inferences (SSIs) of others' behaviors. With some caveats, previous findings suggest that people can spontaneously infer situational cues of behaviors (Duff & Newman, 1997, Lupfer et al., 1995). In a more recent study, Ham and Vonk (2003) found that trait and situational inferences can co-occur for the same behavior. They used an adaptation of the probe recognition task, in which participants were tested not only on traits probes but also on situational probes that had not been seen in the sentences. For example, a situational probe for "Larry lost his job" would be "downsizing." In one study, Ham and Vonk found that participants responded slower to both trait and situational probes compared to control probes, thus suggesting co-occurring activation of both trait and situational inferences.

Another implicit bias assessment paradigm is the "Who said what" paradigm, proposed by Taylor, Fiske, Etcoff, and Ruderman (1978). Originally, this task was designed to measure categorization processes in social perception. Participants were presented with a slide containing a discussion of different topics. The slides featured six speakers, each of which presumably made six statements. In one study (Taylor et al., Study 1), three of the speakers were White and three were Black. After the presentation, participants were asked to match each statement with the picture of the person who had made it. Two measures were assessed: the extent to which participants made within-group errors (e.g., wrongfully attributing a statement made by a White person to another White person) and between-group errors (e.g., wrongfully

attributing a statement made by a White person to a Black person). Participants were more likely to make within compared to between-group errors, thus showing categorization by race. In a more recent study, Coats, Latu and Haydel (2006) adapted the "who said what" paradigm to measure implicit racial bias. In one condition, the majority of the behaviors performed by the Black targets were unfavorable, while the majority of the behaviors performed by White targets were favorable. In a second condition, the favorability was reversed, with most Black targets performing positive behaviors, and most White targets performing negative behaviors. Results showed that participants made more within-group errors in the unfavorable Black condition compared to the unfavorable White condition, thus supporting the idea that categorization was facilitated by exposures to negative stereotypes of Blacks.

Affect Misattribution Procedure. Payne, Cheng, Govorun, and Stewart (2005) developed the Affect Misattribution Procedure (AMP) to measure implicit affective bias towards different groups and targets. Also a sequential priming task, this measure is different than previous ones because it directly asks participants to state their feelings toward ambiguous stimuli. For example, Payne and colleagues (Experiment 6) primed participants with Black and White photos, before presenting them with Chinese ideographs. Participants' task was to evaluate the symbol as being pleasant or unpleasant, using different keys on the computer keyboard. To the extent that participants experience negative affect in response to Black faces, this affect should be misattributed to the ambiguous stimulus that follows it, thus influencing judgments.

Psychophysiological Measures. Several physiological measures have been used to detect implicit bias. These include fMRI to measure amygdala activation (Phelps, O'Connor, Cunnigham, Funayama, Gatenby et. al, 2000), eyeblink startle responses to Black versus White

faces (Amodio, Harmon-Jones, & Devine, 2002), and event-related potentials in response to Black versus White faces (Ito & Caccioppo, 2000). However, probably the most commonly used physiological measure of implicit racial bias is facial electromyography (EMG), which involves measuring the activity of two facial muscles: the corrugator (frowning) muscle and the zygomaticus (smiling) muscle. Greater zygomaticus and lesser corrugator activity indicate positive affect and greater corrugator and lesser zygomaticus activity indicate negative affect. In three studies, Vanman, Paul, Ito, & Miller (1997) found a discrepancy between White participants' self-reported attitudes towards Blacks and their EMG activity: while participants generally reported positive attitudes on explicit measures, their EMG activity reflected negative affective attitudes toward Black targets. Unlike other psychophysiological assessments, this measure has the advantage of capturing both the valence and the intensity of the participant's reaction.

Implicit Bias and Behavior

The existence of implicit bias may only be important to the extent that it causes negative, harmful behaviors toward certain groups. Thus, one important question is whether implicit stereotyping and prejudice are able to predict discriminatory behavior. There are numerous studies that investigate this question and most findings show that implicit measures are more likely to predict subtle, nonverbal behaviors than overt discriminatory behaviors. For example, McConnell and Leibold (2001) found that negative implicit attitudes toward Blacks significantly correlated with a number of negative nonverbal behaviors exhibited towards a Black experimenter compared to a White experimenter, such as speaking and smiling less, displaying more speech errors and hesitations. Participants' implicit attitudes and nonverbal behaviors correlated both when assessed by the Black experimenter and independent judges.

Dovidio, Kawakami, and Gaertner (2002) extended these findings and showed that in an interaction with a Black confederate, White participants' explicit attitudes predicted their verbal friendliness, but not their nonverbal friendliness. However, participants' implicit attitudes predicted confederates' ratings of the participants' nonverbal friendliness. The predictive power of implicit attitudes has been demonstrated not only for racial groups, but also other social groups. For example, Bessenoff and Sherman (2000) showed that participants' implicit associations of fat people predicted how far they chose to sit near a fat person in a potential interaction, such that more negative automatic attitudes were associated with sitting farther from a fat woman's belongings. Surprisingly, biased associations also seem to have an effect on participants' own behavior. For example, Dijksterhuis, Arts, Bargh, and van Knippenberg (2000) found that associating elderly people with forgetfulness predicted participants' own memory impairment, regardless of their own age.

Although implicit measures are best known for predicting subtle, spontaneous outcomes such as nonverbal behavior, some studies have found that they also predict overtly discriminatory behavior. For example, in organizational contexts, a gender IAT predicted budget cuts for racial and ethnic minority organizations (Rudman & Ashmore, 2007) and discriminatory hiring recommendations for Black applicants (Ziegert & Hanges, 2005). Rudman and Glick (2001) also found that associating men, more than women, with agentic characteristics predicted workplace discrimination, such that agentic female, but not male candidates, received negative evaluations when applying for a feminized job. In a more recent study, Latu, Stewart, Myers, Lisco, Estes, and Donahue (in press) found that associating men, more than women, with successful manager characteristics predicted higher salary recommendations for male, but not for female candidates.

Overall, implicit measures seem to consistently show high predictive validity. One important research question is whether implicit measures show significantly higher predictive validity compared to self-report measures. Greenwald, Poehlman, Uhlmann, and Banji (2009) sought to answer this question by conducting a meta-analysis on 184 published and unpublished studies. Studies included in the meta-analysis investigated the relationship between IAT scores and behaviors in different domains such as consumer preference, interracial behavior, personality differences, alcohol and drug use, clinical phenomena, non-racial intergroup behavior, gender and sexual orientation, close relationships and political preferences. Findings showed that the predictive power of implicit versus explicit measure depended on the domain in which the behavior occurred. Both implicit and explicit attitudes reliably predicted behaviors in the domain of consumer and political preference. However, in the more sensitive domain of racial and intergroup interaction, the IAT had significantly greater predictive validity compared to explicit measures.

Another distinction relevant to the predictive power of implicit measures comes from Amodio and Devine (2006), who suggested that implicit stereotyping and implicit evaluation differentially predict behaviors in the domain of interracial interactions. Specifically, implicit stereotyping primarily predicts stereotypic expectations that people have when encountering a target. For example, in two studies, a racial stereotyping IAT significantly predicted the extent to which participants rated an African American target using stereotypic traits, but it did not predict approach-avoidance behaviors. On the contrary, a racial evaluative IAT significantly predicted consumatory behaviors such as seating distance and ratings on a feeling thermometer, but it did not predict stereotypic expectations.

Implicit Bias Reduction

Once researchers reliably documented implicit biases towards different social groups and established that these biases can predict behavior, the focus of research shifted towards identifying successful strategies of reducing implicit intergroup bias. These attempts were initially met with skepticism, as it was assumed that these associations are hard to change because they are automatic, deeply rooted, and outside a person's control (Bargh, 1994; Devine, 1989). However, research conducted in the past twenty years has successfully identified several strategies of reducing implicit bias.

Changing Associations. Given that implicit bias involves associating social groups with negative stereotypic traits or words that are generally negative in valence, one strategy of reducing this negativity involves changing these underlying associations, a strategy that has become "the Holy Grail of implicit race bias research," as Amodio and Mendoza (in press, p. 23) suggest. Techniques of changing these stereotypes include promoting positive counterstereotypes and suppressing existing stereotypic associations.

Counterstereotypes. Exposing participants to counterstereotypes in order to reduce negative intergroup bias was one of the first strategies attempted by researchers. For example, in a series of studies Blair, Ma, and Lenton (2001) had participants engage in a mental imagery task, in which they were instructed to imagine either a strong woman or, in control conditions, a weak woman or a vacation in the Caribbean. Across several implicit gender bias measures, participants who imagined a counterstereotypical strong woman, compared to those engaged in a control imagery task, showed reduced automatic gender stereotyping, such that participants were less likely to associate women, compared to men, with words indicating weakness.

In a later study, Dasgupta and Asgari (2004) found that imagining famous women in high-status positions such as business leaders, scientists, and judges led to being more likely to associate women with leadership characteristics compared to a control condition in which participants were exposed to pictures of flowers. A second study conducted by Dasgupta and Asgari looked at the role of counterstereoyptes in a naturalistic environment, by studying implicit gender attitudes in students enrolled either in a coed college or a women's college, where women occupy most high-powered positions in the college hierarchy. Findings showed that although women enrolled at a women's college and coed college had the same implicit gender attitudes when they enrolled, after a year, women's college students had more positive implicit attitudes compared to coed students, presumably because they have been exposed to positive counterstereotypic models (e.g., female professors). Findings also showed that the frequency of interaction with strong female role models moderated implicit bias reduction effect, such that the more classes students attended, the larger was their reduction in implicit gender bias.

The role of counterstereotypes has not only been studied in the realm of implicit gender bias. In the domain of racial implicit bias, Dasgupta and Greenwald (2001) found less negativity towards Black targets after participants were exposed to an admired Black person compared to a disliked person or a control. These effects were maintained after a 24 hour delay.

Suppression of stereotypes. One strategy to reduce stereotyping is to train participants to negate or suppress current stereotypic associations. Kawakami, Dovidio, Moll, Hermsen, and Russin (2000) designed a negation training, in which participants' task was to repeatedly negate skinhead, elderly, or racial stereotypes. Compared to controls who either did not undergo training or were trained in stereotype maintenance, participants showed reduced

automatic stereotype activation, such that they no longer showed facilitation of categorizing Black and White faces after being primed with Black and White stereotypes respectively. Subsequently, Kawakami, Dovidio, and van Kamp (2005) showed that a variation of this training led to reduced stereotype application in a resume evaluation task, as long as participants were not deliberately trying to avoid being influenced by negation training.

Several criticisms were brought up in response to the Kawakami et al. (2000) negation training. Gawronski, Deutsch, Mbirkou, Seibt, and Strack (2008) argued that the negated stereotype (e.g., “Black NO loud”) can not be stored in memory at an automatic level as such, and thus can actually lead to increased stereotype activation (“Black” was still repeatedly paired with “loud”). Instead, Gawronski and colleagues argue, the reduced stereotype activation effects of the Kawakami training technique are probably due to affirming the counterstereotype, thus forming a new automatic association in participants’ minds (e.g. “Black quiet”). Another issue to consider about negation training is the potentially antagonistic effect of stereotype suppression. Several studies showed that stereotype negation might ironically lead to increased stereotyping (Galinski & Moskowitz, 2000; Macrae, Bodenhausen, Milne, & Jetten, 1994). For example, Macrae and colleagues found that compared to controls, participants who were asked to suppress skinhead stereotypes while constructing a story about a skinhead showed more automatic stereotype activation on a lexical decision task. The authors suggested that this finding was due to a rebound effect – suppressing unwanted thoughts is likely to lead to an increased chance of their reappearance later on. Although the training technique designed by Kawakami and colleagues showed no such rebound effects after 24 hours, it is important to note that stereotype negation may, in certain circumstances, lead to increased stereotype activation.

Changing motivation and goals.

Self-enhancement. One important motivation for social action is the maintenance of a positive self-image. Research shows that such self-enhancement motives can also affect the magnitude of implicit intergroup bias. Sinclair and Kunda (1999) conducted several studies in which non-Black students completed implicit measures of racial stereotyping after receiving either positive or negative feedback from either a Black or a White professional. For example, in one study, participants received videotaped positive or negative evaluations from either a Black or a White manager. Compared to participants who received positive feedback from a White manager, participants who received positive feedback from a Black manager showed reduced negativity towards Blacks. Presumably, these participants were motivated to think highly of a manager who praised them, in order to maintain and validate a positive image of themselves. On the contrary, participants who received negative feedback from a Black target showed high levels of implicit racial bias, as they were motivated to disparage an evaluator who threatened their self-image. Thus, the promotion of a positive self-image does not only affect explicit bias, as predicted by the Social Identity Theory (Tajfel & Turner, 1979), but it also affects intergroup bias at an automatic level.

Social tuning. Individuals' need to belong in social groups and situations often determines them to engage in social tuning – an attempt to adjust attitudes, beliefs, and even memories, to fit with others' perspective, in order to achieve a common ground. Lowery, Hardin, and Sinclair (2001) were interested in how social tuning affects automatic racial attitudes. In past research, social tuning was evident at an explicit prejudice level. For example, participants expressed less explicit racial prejudice when interviewed by a Black compared to a White experimenter (Kinder & Sanders, 1996). However, this finding may be attributed to

social desirability effects, and not necessarily to a true change in prejudiced feelings. Lowery and colleagues challenged the social desirability view and showed that social tuning is also an efficient strategy of reducing bias at an automatic level. In several experiments, White participants showed less automatic negativity towards Black targets in the presence of a Black experimenter compared to a White experimenter.

Later experiments suggested that individuals tend to tune not only to others' mere presence, but also to others' attitudes. In one study, Sinclair, Lowery, Hardin, and Colangelo (2005) revealed that female participants showed reduced automatic prejudice towards Blacks when the experimenter was wearing a t-shirt that expressed social equality message, compared to a neutral t-shirt. Further analyses suggested that this relationship was moderated by how much participants liked the experimenter. Thus, it seems that affiliative motivation, especially when encountering liked others, can reduce automatic bias. This finding was supported in a second experiment conducted by Sinclair and colleagues. In this study, likability was directly manipulated, by having the experimenter behave in either a likable or rude way towards the participant. Findings revealed that participants were less likely to have anti-Black attitudes when a likable experimenter was wearing an anti-racism t-shirt, compared to a neutral t-shirt. This difference did not occur for participants exposed to a rude experimenter. Taken together, these findings suggest that participants adjust their automatic attitudes when exposed to egalitarian messages coming from liked individuals, as a way of social tuning, with the ultimate goal of belonging to social groups.

Changing experience.

Contact with the target. For more than 50 years, social psychologists have been interested in testing the hypothesis that intergroup contact decreases prejudice. These efforts were initiated by Allport (1954), who proposed that contact between different social groups could effectively reduce prejudice if four conditions are satisfied: the groups are of equal status in the contact situation, the groups have common goals, the interaction is cooperative in nature, and the interaction receives the support of authorities. Over 500 studies showed support for the intergroup contact hypothesis, as suggested by Pettigrew and Tropp's recent metaanalysis (2006). Across a large range of groups and contact settings, intergroup contact decreased explicit, self-reported intergroup bias. But does intergroup contact also have the power to reduce implicit intergroup bias? Henry and Hardin (2006) conducted two studies to investigate this hypothesis, and the results were somewhat optimistic. Their findings showed that contact reduced implicit bias, but only for low-status groups. More contact was associated with less implicit bias, but only for Black and not for White participants. The same pattern was found for intergroup contact between Christians and Muslims in Beirut: more contact was associated with less implicit prejudice for Muslims (low-status group), but not for Christians (high-status group). These findings are probably due to low-status deference: low-status groups are motivated to maintain positive (explicit and implicit) views of high status groups. Also, differences in information processing may predict differential automatic stereotypes between the two status groups. Individuals from high-status groups tend to be heuristic (and thus stereotypical) when they encounter individuals from low-status groups. In contrast, low-status individuals are motivated to carefully process information about high-status individuals, and are

thus more mindful and less stereotypical (Fiske, 1993, Keltner, Gruenfeld, and Anderson, 2003).

***Perspective taking*.** Several studies suggested that taking the perspective of stigmatized targets can reduce explicit negative attitudes towards the entire stigmatized group (Batson et al., 1997; Dovidio et al., 2004; Galinsky & Moskowitz, 2000; Vescio, Sechrist, & Paolucci, 2003). The effects of perspective taking seem to also extend to implicit intergroup bias. For example, Galinsky and Moskowitz found that taking the perspective of others, by imagining what the person feels, led to reduced automatic stereotype activation on a lexical decision task.

The context in which the target is perceived. True to the fundamental idea of the power of the situation in social psychology, automatic stereotyping also depends on the context in which the target person is perceived. Wittenbrink, Judd, and Park (2001) found reduced automatic stereotype activation for participants who saw a clip of a Black person in a positive context such as family barbecue, compared to a negative context such as a gang incident. In another study, participants showed more positive associations of Blacks when pictures of Black individuals were presented in a church background compared to a graffiti street corner.

Changing attributions. As early as 1954, Allport suggested that the ways in which we explain others' behaviors have important consequences for intergroup relations. His ideas were later extended and refined by Pettigrew (1979), who was inspired by research on the fundamental attribution error (Gilbert & Malone, 1995; Ross, 1977), which showed that people tend to overly rely on internal, dispositional factors when explaining behaviors, while underestimating situational factors that could explain the same behaviors. Pettigrew extended this attribution tendency to the intergroup domain, and coined the term *the ultimate attribution*

error to describe an ethnocentric pattern of attributions, in which negative behaviors of out-group members, especially if stereotype-consistent, are explained in terms of dispositional or internal factors. This tendency to underestimate situational constraints for out-group members is still pervasive in today's society. For example, people who use this attribution pattern may attribute Black people's criminality rate to their innate aggressiveness, women's professional lack of success to their incompetence, and poor people's unemployment to their unwillingness to get a job. Individual's tendency to engage in the ultimate attribution error has been supported by empirical research. For example, Duncan (1976) found that White participants attributed the same aggressive shove to dispositional factors when the actor was Black, and to situational factors when the actor was White.

The ultimate attribution error is highly relevant for intergroup relations and stereotyping. Consistent with Pettigrew's predictions, this ethnocentric pattern of attribution tends to be enhanced for highly prejudiced individuals (Greenberg & Rosenfield, 1979; Wittenbrink, Gist, & Hilton, 1997) and for groups that have a history of conflict and stereotyped views of each other (see Hewstone, 1990 for a review). The mechanism through which this pattern of attributions perpetuates negative stereotypes is straightforward: when the negative behavior of an out-group members is attributed to internal factors (e.g. He was fired from his job because he is incompetent), while underestimating situational constraints (the declining job market), then it is very likely that the negative stereotypes about the out-group are reinforced and perpetuated. For example, Brescoll and Uhlmann (2008) showed that women who express anger are accorded lower status than men, and this relationship was mediated by internal attributions of anger expression (e.g., women express anger because they are out of control, men express anger because their situation is aggravating).

Existing research seems to indirectly suggest that situational attributions for out-group members' behaviors may underlie implicit bias reduction. Stewart et al. (2010) designed a novel training technique aimed at reducing racial stereotyping by changing attributions – the basic pillars on which stereotyping stands. This technique, called Situational Attribution Training, had participants replace dispositional, stereotypic explanations with situational explanations of undesirable behaviors of Black men. The findings of two experiments showed that repeated training in making situational attributions for negative stereotypic behaviors of out-group members led to decreased automatic stereotype activation. Across two experiments, participants were randomly assigned to one of two conditions. In the Situational Attribution Training condition participants repeatedly chose situational over dispositional explanations of stereotypic negative behaviors presumably performed by Black males. Depending on the experiment, control participants either did nothing (No Training Control condition) or completed a similar task to Situational Attribution Training, but instead of choosing situational over dispositional explanations, they had to count the number of nouns and verbs in the behavioral sentences (Grammar Training Control condition). Finally, all participants completed the Person Categorization Task (Blair & Banaji, 1996) as measure of automatic stereotype activation. This task required participants to categorize quickly White and Black photos after being primed with positive and negative traits that were stereotypic either of Blacks or of Whites. Across two studies, control participants who did not undergo Situational Attribution Training categorized Black photos significantly faster than White photos after being primed with a Black negative stereotype, thus showing automatic stereotype activation of Blacks. However, experimental participants who completed Situational Attribution Training no longer showed facilitation of categorizing Black faces compared to White faces after being primed with negative Black

stereotypic traits. In other words, the two studies showed consistent evidence that extensive training in making situational attributions for negative behaviors decreases the automatic association between Black negative stereotypes and Black targets. In Allport's words, the Stewart et al. training technique "strenuously disciplined" individuals' tendency to engage in the ultimate attribution error and consistent with Allport's suggestion, this strategy led to reduced intergroup bias.

Although Stewart et al. (2010) was the first to document the causal relationship between situational attributions and stereotype reduction, another study came close to accomplishing that goal. Designed as a statistical training, Schaller, Asp, Rosell, and Heim (1996) taught participants the logic behind analysis of covariance and how it applies to our explanation of everyday outcomes. In one condition, participants learned how to take into account situational factors that may determine the racial achievement gap on standardized tests. Compared to controls, participants who underwent training were less likely to show stereotyping of minimal groups in a subsequent task. However it should be noted that situational attributions were only one of the components of the covariance training. In addition, the study did not document the effects of training on implicit stereotypes, or on racial stereotypes, as did Stewart and colleagues in the Situational Attribution Training studies.

Automaticity and Social Inferences

Automatic and Controlled Processes in Social Cognition

Any process, be it driving a car or making an impression of a new person we meet, usually ranges between two poles: automatic and controlled. In controlled processes, the person is in control of their thoughts, feelings, or behaviors. In automatic processes, this control is

taken over by the environment, with little input from the person. According to Bargh (1994), there are four characteristics that make a process automatic: unawareness, lack of intention, efficiency, and lack of control. A person can be unaware of a certain mental process in several ways. First, a person can be unaware of the presence of a certain stimulus, as is the case of subliminal perception and priming. For example, Bargh and Pietromanco (1982) showed that the more participants were primed subliminally with words related to hostility, the more likely they were to interpret an ambiguous behavior as hostile. Second, even if a person is aware of a stimulus, they can be unaware of how the stimulus influences certain mental processes, as is the case in priming studies. For example, in the lexical decision task (Wittenbrink et al., 1997) participants are unaware that the race of the photo they see on the screen may influence the speed with which they decide whether subsequently presented (stereotypic) traits are words or non-words. Finally, in automatic processes, individuals may be unaware of what exactly determined a certain judgment or feeling, as is the case in the Affect Misattribution Procedure (Payne et al., 2005), in which subjective judgments of ambiguous symbols are, without the participants' awareness, influenced by their feelings towards different racial groups to which they have been exposed.

A second characteristic of automatic processes is their decreased intentionality, the degree to which one is in control when it comes to initiating a certain process. A prime example of unintentionality is automatic stereotype activation. For example, Devine's famous studies (1989) were among the first ones to show that Black stereotypes are unawarely activated by exposure to African American traits, such that participants primed with these traits were more likely to interpret ambiguous behaviors as hostile compared to participants who were not primed.

Efficiency is the third characteristic of automatic processes. According to Bargh, a process is efficient to the extent to which it does not require a great deal of attentional resources. For example, trait (dispositional) inferences in person perception are generally more cognitively efficient compared to situational inferences, which require cognitive resources to be completed.

Finally, automatic processes are characterized by a relative lack of control, the extent to which the person is in control of ending a particular process. According to several researchers, motivations play an important part in moderating the degree of control that a person has over a mental process: motivations to control prejudice (Bargh, 1994; Devine, 1989), need for cognition (Bargh, 1994; Fiske, Lin, & Neuberg, 1999), or personal need for structure (Fiske et al., 1999). Higher levels of these motivations are likely to increase control over certain processes.

Traditionally, mental processes have been viewed as either automatic or controlled. However, more recently, researchers have emphasized the need to view automaticity as a continuum. As such, a process can never be purely automatic or controlled. Instead, it varies in the degree to which it has automatic and controlled components. For example, stereotype activation has both automatic components (e.g., racial stereotypes are often unawarely activated in the presence of a target person; Devine, 1989) and controlled components (e.g., stereotypes require some attentional resources to be activated; Gilbert & Hixon, 1991). Related to person perception, Fiske et al. (1999) proposed a continuum model, with category-based (automatic) processes at one end of the continuum and individuating (controlled) processes at the other end. Fiske and her colleagues maintain that perceivers give priority to category-based processes because they are quicker and more efficient than individuating processes. However, several

factors, such as motivation and availability of information can attenuate this tendency, pushing processes towards the more controlled end of this continuum.

Assessing Automaticity in Social Cognition

Cognitive load. A common way to study the degree to which a process is automatic is to limit participants' cognitive resources while performing the task of interest, by having them engage in a competing cognitive task (Navon & Gopher, 1979). If the process is automatic, then the competing cognitive task should not disrupt performance on the primary task. If the process is not automatized, the cognitive task should disrupt individuals' performance, suggesting that the process requires some attentional resources. Several cognitive load tasks have been employed in the cognitive and social literature. These include monitoring flashing lights (Osterhaus & Brock, 1970) or Xs (Petty, Wells, & Brock, 1976) on a screen, with distraction level being manipulated by increasing the rate of flashing stimuli. Another cognitive load task involves engaging participants in competing listening tasks, such as tracking musical tone changes (Skitka, Mullen, Griffin, Hutchinson & Chamberlin, 2002), listening to music and keeping track of the number of songs played (Lalwani, 2009), indicating each time a particular sequence of tones was played (Gilbert, Gill, & Wilson, 2002), and hearing and repeating a story recorded on a tape (Mikulincer, Birnbaum, Woddis, & Nachmias, 2000). Another line of cognitive business tasks simply restricted participants' window of responding in a decision-making task: Deutsch, Kordts-Freudinger, Gawronski, and Strack (2009) used the Affect Missatribution Procedure and restricted participants' response window to below 600ms (high load) or above 600ms (low load). Gilbert and Gill (2000) also used time pressure in decision making, by allowing participants either 2 seconds (high load) or 10 seconds (low load) to make a decision of whether they had seen a stimulus before or not.

By far, the most popular cognitive load task in social cognition has been the Gilbert and Hixon (1991) procedure of having participants memorize numbers while performing the task of interest. In the high cognitive load conditions, participants are presented with five to eight digit numbers at the beginning of the task, and asked to memorize the digits, because they would have to recall them at the end of the task. In the low cognitive load condition, participants memorize and recall one digit numbers. Presumably, participants silently rehearse the digits while completing the main task, with more digits being associated with more cognitive business. Although this is the main paradigm, several variations of this paradigm exist in the social cognition literature. A summary of cognitive load manipulations using the Gilbert and Hixon paradigm is presented in Table 1.

Table 1. Cognitive Load Manipulations in the Literature

Study	Main findings	Task	Manipulation of Cognitive Load	Cutoff for elimination	Reaction Times
Osborne and Gilbert, 1990 (unpublished manuscript)	Participants responded slower to probes 2 minutes after load, so they were probably rehearsing the number		20s to memorize 8-digit number	N/A	N/A
Gilbert and Hixon, 1991	Cognitive business impairs stereotype activation but it stimulates stereotype application	Word-fragment completion task	Memorize 8-digit number and recall it at the end of the task	4 or more errors	N/A
Sherman et al., 1998	Under high cognitive load, stereotype-inconsistent information receives greater attention	Impression formation task – read sentences, reading time recorded	Memorize 8-digit number	Participants not eliminated because they did not make any errors	N/A
Pendry and Macrae, 1999		Impression formation task	20s to memorize 8-digit number and recall at the end of task	N/A	N/A
Sherman and Frost, 2000	Under high cognitive load, recall is better for stereotype-consistent info (compared to inconsistent) and recognition is better for stereotype-inconsistent info	Impression formation task – test recall and recognition	Memorize 8-digit number and recall it at the end of the task	N/A	N/A
Koole et al., 2001	Implicit self-esteem predicts positive self-evaluations under high but not low cog load	Trait rating (“me” or “not me”)	High load: memorize 8-digit number; Low load: memorize 1-digit number; Recall at the end of the task	N/A	N/A

Bodner and Stalinski, 2008	Priming in the lexical decision tasl is automatic, and would endure under high cognitive load.	Lexical Decision Task	Memorize 8 digit number, recognition tested after each trial	N/A	RT no load < RT load - no interaction with priming – same effects under load and no load -> automatic
Van den Bos et al., 2006	People are more satisfied with advantageous inequity under high versus low cognitive load	Reading about a hypothetical situation and rating their satisfaction of the outcome	High load: strings of 8 symbols @ * % # ? $ + and, presented for 25 s, and then asked to recall at the end of the experimental procedure; low load pp were just given one symbol @.	N/A	Satisfaction with advantageous inequity higher under high versus low load
Stewart et al., 2003	Participants spontaneously stereotype, regardless of level of prejudce	Probe Recognition Task	High load: memorize 5-digit number; low load – 1-digit number; recall after each trial	N/A	N/A - only accuracy data
Wigboldus et al., 2004	Under high load: STIs more likely for ster consistent vs inconsistent behaviors. Under no load: no difference	Probe Recognition Task	Study 1: memorize 8-digit number at the beginning of task, recall at the end Study 2: high load – 5-digit number; low load – 1-digit number; recall at the end of each trial.	4 or more; ran analyses with and without those participant – no difference – so reported those with error pp in	High load: ~ 1030ms Low load: ~ 1010ms

Notes. Low load and no load seem to be interchangeable. Low load (one-digit number) seems to be used more in trial-by-trial recall tasks than no load, probably to maintain load conditions somewhat equivalent.

Cognitive load and implicit bias. As early as Allport's initial writings on prejudice (1954), social psychologists have agreed that social categorization and stereotyping serve the function of organizing the social world into clear, well-known categories, thus increasing cognitive economy. As Gilbert and Hixon wrote, "A stereotype is a sluggard's best friend" (1991, pp. 509). The consensus has been that if stereotypes serve such cognitive economy purposes, reducing a perceiver's cognitive resources should increase the tendency to stereotype. Several studies have shown that when individuals are short on attentional resources, they are more likely to rely on stereotypes and less likely to use individual information about the target person (Bodenhausen, 1990). However, later studies (Gilbert & Hixon, 1991; Spencer, Fein, Wolfe, Fong, & Dunn, 1998) found evidence of less stereotyping under conditions of high compared to low cognitive load. Their findings suggest that cognitive business differentially impacts stereotype activation versus stereotype application. Because stereotype activation is not a fully automatic process and requires some cognitive resources to be completed, cognitive load disrupted the activation of stereotypes, thus leading cognitively busy participants to make less stereotyping word completions compared to participants who were not busy. However, consistent with previous findings, the same participants showed more stereotype application under high compared to low cognitive load.

Spontaneous Inferences

One interest of researchers is how people infer the causes of other people's behaviors. If we hear that Rick lost his job, we can infer that there's something about Rick that got him fired (he is not hard-working) or that there's something about the situation that determined Rick to

lose his job (the economy is in recession). Earlier social inference theories (e.g., Gilbert & Malone, 1995) proposed that individuals initially attribute the causes of a behavior to internal or trait factors. This tendency was called the fundamental attribution error or correspondence bias. According to Gilbert and Malone, the inference process has three stages. First, the perceiver categorizes the behavior. Second, the perceiver draws an internal, dispositional inference of that behavior. A third stage involves correcting for the dispositional inference to take into account situational factors. This stage is optional and it only occurs if the perceiver has sufficient cognitive resources and motivation to make the correction. One implication of this model is that the second stage of drawing the dispositional inference is somewhat automatic, occurring without a person's effort and awareness. However, the research paradigm used to assess the way individuals make social inferences does not seem to adequately assess the automaticity of these inferences. For example, in Gilbert, Pelham, and Krull (1988), participants are presented with a videotape of a woman behaving anxiously. Afterwards, participants are asked whether the woman behaved that way because she was nervous or because the situation was anxiety provoking. This paradigm forces participants to make an intentional inference, and thus is not adequate for assessing automatic social inferences.

A more appropriate paradigm for investigating the way people automatically infer the causes of others' behaviors comes from the spontaneous inference literature (see Uleman, Newman, & Moskowitz, 1996b for a review), using the probe recognition task (McKoon & Ratcliff, 1986). Unlike classic attributions, these spontaneous inferences occur without an impression formation goal in mind. Instead, they are guided by chronically accessible constructs. Because they are spontaneous, no cognitive effort is required in drawing these inferences. For example, upon hearing that Rick lost his job, a perceiver may spontaneously

think “incompetent,” thus making a spontaneous trait inference (STI). There is consistent evidence (Lupfer, et al., 1995; Stewart et al., 2003; Uleman, et al., 1996a) that perceivers make STIs without effort and without being motivated to do so. Some research also suggests that people may also make spontaneous situational inferences (SSI) when perceiving a behavior (Duff & Newman, 1997, Lupfer et al., 1995), and that STIs and SSIs can co-occur for the same behavior (Ham & Vonk, 2003).

Bargh (1994) claimed that spontaneous inferences are almost fully automatic. Previous research partially supports this claim. For example, Wigboldus and colleagues (2004) showed that cognitively loading participants while completing the probe recognition task led to impairments of STIs for stereotype inconsistent behaviors. Thus, it may be that spontaneous inferences of stereotypic behaviors are more likely to lie towards the automatic end of the continuum compared to those of nonstereotypic behaviors.

Individual Differences and Implicit Bias

There are several individual differences that moderate individuals’ implicit bias. These moderators include explicit prejudice, motivations to control prejudice, and cognitive complexity variables such as need for cognition and personal need for structure.

Explicit Prejudice

How does explicit prejudice relate to implicit bias? There has been a large debate in the literature about the relationship between explicit and implicit attitudes. While some researchers suggest there is not enough evidence to claim that explicit and implicit attitudes are conceptually different (Fazio & Olson, 2003; Gawronski et al., 2007), others support the idea of

independent constructs (Nosek & Smyth, 2007). Research findings also seem to maintain this controversy, with some explicit and implicit measures being significantly related to each other and others being unrelated. For example, self-report measures of hostile sexism did not correlate with several gender IATs, but benevolent sexism significantly correlated with some implicit associations, such as associating men, more than women, with high status words and agentic traits (Rudman & Kilianski, 2000). In other domains, Nosek and Smyth found that self-report ratings significantly correlated with implicit ratings for attitudes towards gays, but not for attitudes towards Blacks. Overall, findings seem to support the idea of construct independence. Using a multitrait-multimethod design, Nosek and Smyth found that a dual-attitude factor model of implicit and explicit attitudes fit significantly better than a single-attitude factor model, thus supporting the idea of distinct attitude constructs.

Motivations to Control Prejudice

It is widely accepted that some societal norms discourage the expression of racial bias and that people are less likely to express their prejudice beliefs in order to conform to these norms. Researchers also proposed that motivations to be fair might stem not only from external sources such as social norms, but also from internal sources, such as personal, internalized beliefs about social equality and fairness. For example, Plant and Devine (1998) proposed that there are two sources of motivation that determine people to control their prejudiced responses to Black individuals. First, individuals may try to respond without prejudice because it is personally important for them to foster non-prejudiced beliefs, thus relying on internal motivations to control prejudice (IMS). Second, people may try to control their prejudiced responses because they want to avoid being seen negatively by others – a concept named external motivation to control prejudice (EMS). These two concepts are independent of each

other, and individuals may be characterized by high levels of either or both motivations. Across several studies, Plant and Devine developed and established the convergent and discriminant reliability of separate scales that measure IMS and EMS. Later research investigated how the combination of these motivations was related to the level of implicit racial bias. In several studies, Devine, Plant, Amodio, Harmon-Jones, and Vance (2002) found that individuals high in IMS and low in EMS displayed lower implicit racial bias than did individuals characterized by any other IMS/EMS combination. These effects were maintained for several implicit bias measures, such as a lexical decision task and the IAT (Greenwald et al., 1998). Participants high in IMS showed less implicit racial bias because their non-prejudiced beliefs became internalized over time, thus leading to implicit positive attitudes toward Blacks. Those also low in EMS were less likely to be concerned with how they looked to others and more likely to act and feel according to their non-prejudiced beliefs. Importantly, participants characterized by this motivational profile maintained low levels of implicit racial bias even under conditions of cognitive business, thus suggesting that they are not necessarily better able to control their prejudice responses, but they have less stronger associations between Blacks and negativity.

Cognitive Complexity

When individuals have the cognitive capacity to process information, they are more likely to engage in controlled, individualized processing of information about outgroup members. This propensity to process information depends, to some extent, on individual differences, which include NFC and PNS.

Need for cognition is individuals' tendency to engage in and enjoy thinking (Cacioppo & Petty, 1982). Cacioppo and Petty tested the reliability and validity of a 45-item scale to

measure NFC, which was later adjusted by Cacioppo, Petty, and Kao (1984) into a short version consisting of 18 items. Individuals high in NFC typically show greater cognitive complexity, as they are more likely to seek and enjoy tasks that are cognitively challenging. Relatedly, these individuals are also less likely to rely on heuristics and simple cognitive structures such as schemas when processing social information. Thus, NFC is significantly related to stereotyping, consistent with several dual processing models which maintain that cognitive complexity is a key variable that moderates the extent to which a mental process is automatic or controlled. For example, Bargh (1994) proposed that individuals high in NFC are more likely to control their mental processes, and thus show less automatic processing. Similarly, Fiske et al. (1999) proposed in the continuum model that people are more likely to situate themselves at the individuating end of the processing continuum when they are high in NFC. The hypothesis that individuals low in NFC are more likely to use stereotypes was supported by several studies (Crawford & Skowronski, 1998), which found that individuals low in NFC remembered more stereotype-consistent information compared to those high in NFC.

A second important concept is that of personal need for structure. Neuberg and Newsom (1993) proposed that "people meaningfully differ in the extent to which they are dispositionally motivated to cognitively structure their worlds in simple, unambiguous ways" (p. 114), as way to reduce cognitive overload. Thus, they conceptualized "need for structure" as the dispositional need to simplify and structure one's environment in simple ways. To measure individuals' tendency to prefer and use simple cognitive structures, Neuberg and Newsom developed the Personal Need for Structure (PNS) scale, which has been shown to possess sufficient reliability (Cronbach α = .77) and convergent and discriminant validity. The tendency to structure the social world is related to the use of cognitive structures such as schemas, as a way to reduce the

complexity of one's environment. It follows that high need for structure should be associated with greater stereotyping of others. Theoretically, this proposition is supported by dual process models of automatic and controlled processing. For example, Fiske and her colleagues (1999) proposed that individuals who are high in need for structure are more likely to categorize others as group members compared to their low in need for structure counterparts. This hypothesis was supported by Neuberg and Newsom who found that individuals high in need for structure showed greater gender stereotyping, such that they were more likely to attribute negative stereotypic traits to female compared to male targets.

Importantly, although past research has investigated how NFC and PNS relate to implicit bias and stereotyping, so far no studies have successfully documented how these personality variables moderate participants' reactions to trainings designed to reduce prejudice and stereotyping.

The Present Research

In the current studies I focused on the effectiveness of an implicit bias reduction technique, Situational Attribution Training (Stewart et el., 2010). In short, this training technique focused on undoing the fundamental attribution error (Pettigrew, 1979), the tendency to attribute negative behaviors of outgroup members to dispositional factors while underestimating the role of situational factors. The fundamental attribution error is one of the pillars on which stereotyping stands, because attributing negative behaviors of outgroup members to stable, internal traits tends to strengthen negative outgroup stereotypes. The Situational Attribution Training technique was designed to circumvent this attributional tendency by teaching White participants to attribute negative stereotypic behaviors of Black

men to situational causes. Initial findings showed that participants trained in making situational attributions were less likely to show automatic activation of negative Black stereotypes compared to control participants.

The goal of the current two studies was to develop a model of Situational Attribution Training, by investigating the social-cognitive mechanisms through which training leads to decreased stereotype activation. I designed the current research to answer three main questions. First, I was interested in the effects of training on situational inferences; second, I investigated the cognitive costs of training; last, I was interested in the role of individual differences in stereotype reduction following training. Taken together, the two current studies were designed to investigate how Situational Attribution Training succeeds in reducing automatic stereotyping and for whom it is most effective.

Study 1

First, what is Situational Attribution Training changing? Previous research showed that it reduces the association between Blacks and negative stereotypic traits; however, little is known about the effects of training on the tendency to make situational inferences of negative behaviors performed by Blacks, which may explain the effects of training on reduced stereotype activation.

To address this issue, I turned to the spontaneous inference literature. Spontaneous inference theories look at how people infer the causes of other's behaviors without awareness, intention, and expenditure of many cognitive resources. It is generally agreed upon that people spontaneously make trait inferences (STIs) about others' behaviors (Lupfer et al., 1995; Stewart et al., 2003; Uleman et al., 1996a), especially for stereotype-consistent behaviors (Wigboldus et

al., 2004). The most common way to study STIs is the probe recognition task (McKoon & Ratcliff, 1986). Construed as a memory task, this paradigm involves presenting people with several behavioral sentences (e.g., “Larry lost his job.”). Following the presentation of each sentence, participants are presented with probe words and decide as quickly and accurately as possible whether the word appeared in the sentence or not. Of interest is the trait probe – a negative trait that was implied but not present in the behavioral sentence (e.g., “incompetent”). Participants who spontaneously activated this trait while reading the sentence had more difficulty rejecting this probe, as suggested by both accuracy data – being more likely to incorrectly identify that a trait appeared in the sentence (Lupfer et al., 1995; Stewart et al, 2003; Uleman et al, 1996a) and reaction time data– being slower to correctly reject a trait (Uleman et al, 1996a, Wigboldus et al., 2004).

In a more recent study, Ham and Vonk (2003) adapted the probe recognition task to investigate the extent to which people also make situational inferences of others’ behaviors. Instead of being presented with trait probes following a behavioral sentence, participants quickly decided whether a situational probe had appeared in the sentence (e.g., “downsizing” for “Larry lost his job”). In one study, Ham and Vonk found that participants were slower to reject both trait and situational probes compared to control probes, suggesting automatic activation of trait and situational inferences.

In the first study, I used Ham and Vonk’s (2003) adaptation of the probe recognition task to investigate how participants’ SSIs are affected by Situational Attribution Training. Previous findings using this training technique showed that participants are less likely to associate Black faces with negative stereotypic traits. The current study allowed us to understand whether this trait effect was accounted for by an increased tendency to infer

situational causes of behavior spontaneously. Following Situational Attribution Training, participants completed an adaptation of the probe recognition task designed to measure the extent to which participants make SSIs after being exposed to negative, stereotypic behaviors presumably performed by Black men. If Situational Attribution Training increases the tendency to attribute negative stereotypic behaviors to situational factors, participants should be slower to reject the situational versus control probes after completing Situational Attribution Training, but not control training. These effects may also be visible for accuracy data, as participants may be more likely to incorrectly decide that a situational probe appeared in the sentence compared to a control probe after Situational Attribution Training, but not control training.

Study 2

How automatic are the effects of Situational Attribution Training? Compared to controls, trained participants in Stewart et al. (2010) were less likely to associate negative Black stereotypic traits with photos of Black men. Was this result obtained because training succeeded in having participants automatically take into consideration previously underestimated situational variables? In order to answer this question, I revisited the theory behind how people make attributions of others' behaviors. Gilbert and Malone (1995) proposed that people tend to make dispositional inferences about others' behaviors automatically, especially when they are trying to understand people and not situations (Krull & Erikson, 1995). Later, they may consciously attempt to correct these dispositional attributions in order to take into account the situation, but only if they have sufficient cognitive resources to do so. I hypothesize that Situational Attribution Training is automatizing the tendency to correct for the initial tendency to make trait inferences, thus weakening the association between Blacks and negative stereotypic traits. If these hypothesized corrections are indeed becoming automatic, I expect the

stereotyping reduction effects of Situational Attribution Training to be maintained in conditions involving both high and low cognitive resources.

As commonly seen in the literature, I limited participants' cognitive resources while performing the dependent task, by having them engage in a competing cognitive task. This competing task should not disrupt performance on the primary task if the process of correcting for trait inferences is automatic. However, no process is purely automatic or controlled; instead automaticity lies on a continuum (Bargh, 1994). Thus, it is possible that training does not make the corrections fully automatic; instead, training may increase the degree to which these corrections are automatic. In Study 2, following training, half of the participants completed the measure of automatic stereotype activation under conditions of cognitive load and the other half under conditions of no cognitive load.

Unlike previous studies that investigated the stereotyping reduction effects of Situational Attribution Training (Stewart et al., 2010), in the current study I used the probe recognition task as a measure of automatic stereotype activation. This task was preferred because it allows for the investigation of the effects of cognitive business on implicit bias. Unlike the person categorization task, which was used in Stewart et al., the probe recognition task has been used successfully in the past in conjunction with cognitive load tasks (Stewart et al., 2003; Wigboldus et al., 2003). At the same time, using the probe recognition task was an opportunity to replicate the stereotyping reduction effects of training with a different paradigm.

Although the person categorization task and other similar sequential priming tasks are primarily used to measure implicit bias, it should be noted that the probe recognition task is also used sometimes. All these tasks use response times to measure implicit bias. In sequential

priming tasks, faster reaction times to stereotype consistent pairings (e.g., Black photo and negative trait) compared to inconsistent ones suggest automatic stereotype activation. In the probe recognition task, slower reaction times to rejecting trait probes compared to control probes for negative stereotypic behaviors suggest automatic activation of negative trait inferences. Thus, compared to sequential priming tasks, the probe recognition task is a more indirect measure of automatic stereotype activation – it primarily measures the extent to which participants make automatic trait inferences for negative stereotypic behaviors.

In the current study, participants in the control condition were expected to show automatic activation of negative stereotypic trait inferences, such that they would be slower to reject the trait versus the control probes for negative Black stereotypic behaviors. However, if the effects of Situational Attribution Training replicate in the probe recognition task paradigm, participants who complete Situational Attribution Training should not show spontaneous activation of negative Black-stereotypic traits, such that they would not be slower to reject the trait versus the control probes for negative Black stereotypic behaviors. In addition, if participants' tendency to make situational inferences has become automatic after training, there should be a decrease in stereotype activation regardless of whether they are cognitively loaded or not.

Last, who is most likely to benefit from training? Despite an abundance of studies that investigated the effectiveness of implicit bias reduction techniques, there are very few studies that investigated the moderating effects of individual differences on prejudice reduction (for a review, see Levy, 1999). Another goal of Study 2 was to investigate individual differences as moderators of the effects of training on automatic stereotype reduction. These individual

differences included explicit prejudice, motivations to control prejudice, and cognitive complexity (NFC and PNS) and were assessed before the training phase through several scales.

Are participants high in explicit prejudice less likely to respond positively to stereotyping reduction interventions? Surprisingly few studies have investigated this question. For example, Monteith (1993) induced participants to believe they discriminated against a gay law school applicant based on his sexual orientation. Participants who were initially low in prejudice showed more negative emotional reactions to this dissonance and less prejudice on a subsequent task compared to participants who were initially high in prejudice. Similarly, I propose that low-prejudice individuals would show more pronounced stereotype reduction effects after undergoing Situational Attribution Training. In other words, training would work best for participants who were initially low in explicit prejudice. This finding would also be consistent with findings that show that low-prejudice individuals are less likely to engage in the ultimate attribution error (Greenberg & Rosenfield, 1979; Wittenbrink et al., 1997). Thus, these participants would also show more automatic stereotype reduction after a training aimed at reducing this biased attributional pattern.

A second individual difference variable that may influence reactions to stereotyping reduction interventions is motivation to control prejudice, as suggested by previous research. For example, Allen, Sherman, and Klauer (2010) found reduced implicit bias when a Black target was presented in a positive (church) context compared to a negative (prison) context, and this effect was moderated by motivation to control prejudice (Dunton and Fazio, 1997). Specifically, motivation did not matter when the target was presented in a negative context, such that implicit bias was high at all levels of motivation. However, when the target was presented in a church context, higher motivation was associated with less bias. In the current

study, I investigated the role of internal and external motivations to control prejudice (IMS, EMS, Plant and Devine, 1998) in moderating responses to Situational Attribution Training. Internal motivation describes individuals' tendency to control their biased responses because it is personally important for them to foster non-prejudiced beliefs. External motivation stems from individuals' desire to conform to social norms that sanction the expression of prejudiced beliefs. Previous research (Devine et al., 2002) found that individuals who are at the same time high in IMS and low in EMS show the least amount of implicit racial bias. These findings were replicated for implicit gender stereotyping by Latu and colleagues (in press). In the current studies I propose that high IMS/low EMS individuals would also show the largest amount of stereotype reduction following Situational Attribution Training. Individuals who are high in IMS are intrinsically motivated to reduce stereotyping, and would thus be more likely to internalize the stereotype reduction strategy of considering situational factors in explaining Blacks' negative behaviors. Low EMS individuals were also expected to show significant implicit bias reduction. Although high EMS participants may be likely to respond to experimental cues that promote anti-bias norms, it is unlikely that this motivation would play a role in stereotype reduction following training, because such external sources of motivation may not affect automatic responses.

Individual differences in cognitive complexity, such as need for cognition – the tendency to engage in and enjoy thinking (Cacioppo & Petty, 1982) and personal need for structure – the tendency to prefer simple cognitive structures (Newberg & Newsom, 1993) may also moderate the effects of Situational Attribution Training on automatic stereotype reduction. Despite the theoretical and applied importance of this hypothesis, very few studies have successfully documented the moderating role of NFC or PNS on participants' reactions to bias

reduction techniques. For example, Schaller et al. (1996) found that PNS did not moderate the effects of a prejudice reduction technique on participants' stereotyping of minimal groups. A different, but related concept of preference for consistency (Cialdini, Trost, & Newsom, 1995) received some attention in prejudice reduction research. Preference for consistency refers to the individual tendency to be and appear consistent with one's own responses and the desire that others be consistent. Heitland and Bohner (2009) found that individuals high in need for consistency showed less prejudice following a prejudice reduction intervention compared to their low need in consistency counterparts. In the current research I investigated the role of participants' initial level of cognitive complexity in their reaction to Situational Attribution Training. This training technique aimed at teaching participants to automatically take into account complex factors that may determine negative behaviors performed by Black men. Thus, individuals who are naturally inclined to prefer cognitive challenges may be likely to complete correctly and benefit from training. I propose that high cognitive complexity would be associated with more automatic stereotype reduction following Situational Attribution Training. Specifically, individuals high in NFC and those low in need for structure would benefit most from Situational Attribution Training. However, these individuals may only be able to process and internalize the additional situational information (thus correcting for the trait inferences) if they have sufficient cognitive resources to do so. Thus, the positive effects of training for participants high in cognitive complexity may only be visible under conditions of low, but not high cognitive load.

Hypotheses Overview

Hypothesis 1. In Study 1, Situational Attribution Training would increase the tendency to make spontaneous situational inferences for negative stereotypic behaviors performed by

African American men. Statistically, participants would be slower to reject situational compared to control probes after completing Situational Attribution Training. This effect would be inexistent or less strong for participants in the control condition.

The overarching goal of Study 2 was to investigate the effects of Situational Attribution Training on automatic stereotype activation. In the probe recognition task, which was used as a dependent measure, automatic stereotype activation is equivalent to the activation of STIs for negative stereotypic traits. To show STI activation, participants should be slower to reject the trait compared to the control probes after being exposed to negative Black stereotypic behaviors. A stereotyping score was obtained by subtracting the average response time to control probes from the average response time to trait probes, such that positive stereotyping scores suggest automatic stereotype activation, and negative stereotyping scores suggest a lack of automatic stereotype activation. The dependent variable used in all analyses was this stereotyping score.

Hypothesis 2. Participants who complete Situational Attribution Training would show reduced stereotype activation, as suggested by decreased STI activation for negative stereotypic traits. There would be a significant effect of training condition on the stereotyping score, such that training participants' stereotyping scores would be significantly lower than those of control participants.

Hypothesis 3. If the effects of Situational Attribution Training have become automatic, participants would show reduced automatic stereotype activation regardless of whether they were cognitively loaded or not when completing the probe recognition task. Statistically, there would be a significant effect of training condition on stereotyping scores and this effect would not be qualified by a significant interaction between training and load condition, such that

stereotyping scores would be lower for training compared to control participants regardless of load condition.

Hypothesis 4. Individual differences such as explicit prejudice, motivations to control prejudice, NFC and PNS would moderate the effects of training on automatic stereotype activation.

Hypothesis 4a. Individuals who were initially low in explicit prejudice would show the most benefits from Situational Attribution Training. There would be a significant interaction between explicit prejudice and training condition, such that the relationship between prejudice and stereotyping would be significantly more positive in the training, but not in the control condition.

Hypothesis 4b. Participants high in IMS and low in EMS would show the most automatic stereotype reduction after Situational Attribution Training. I expected a significant interaction between IMS, EMS, and training condition. For participants who completed training, there would be a significant interaction between IMS and EMS, such that higher IMS would be associated with less stereotyping at low levels of EMS, but not at high levels of EMS. This interaction between IMS and EMS would not be significant for control participants.

Hypothesis 4c. Stereotype activation would decrease after Situational Attribution Training, and these effects would be especially large for high NFC individuals. Also, I expected this effect to occur only under no load, but not under load conditions. Statistically, I expected a significant interaction between training condition, load condition and NFC. The relation between NFC and stereotyping would be significantly more negative for individuals in the no cognitive load-training condition, relative to the no cognitive load-control condition. No such effects should be seen in all conditions in which there is load.

Hypothesis 4d. I expected a significant interaction between training condition, load condition, and PNS. The relation between PNS and stereotyping would be significantly more positive for no load-training individuals compared to no load–control individuals. There would be no such effects in the load conditions.

CHAPTER 2: METHOD

Two experiments were conducted to investigate the mechanisms and moderators of Situational Attribution Training. In Study 1, I investigated the effects of training on spontaneous situational inferences. Study 2 was more extensive and investigated the effect of training on spontaneous trait inferences, the extent to which the effects of training have become automatic, and whether individual differences such as explicit prejudice, motivations to reduce prejudice, and cognitive complexity moderate responses to training.

Study 1

Participants and Design

Eighty-one non-African American students (67 White, 11 Asian American, 1 Latino/a, 1 other) enrolled in Introductory Psychology courses participated in the experiment, as one means to fulfill a course requirement. Twenty-four participants were male and 72 were native English speakers. Participants were randomly assigned to either the Situational Attribution Training or the Grammar Control condition. Afterwards, all participants completed an adaptation of the probe recognition task, designed to measure the degree of activation of spontaneous situational inferences.

Procedure and Materials

Phase 1: Situational Attribution Training. Upon coming into the lab, participants were randomly assigned to either the Situational Attribution Training condition or to the Grammar Control condition. In the Situational Attribution Training condition participants were informed that they would be taking part in a study that investigated how people explain others' behaviors. The experimenter explained and exemplified the difference between dispositional and situational explanations and informed participants that they were randomly assigned to a condition in which they would make situational attributions of others' behaviors. These instructions were repeated once participants started the computer self-paced program. The program instructions also informed them that they were randomly assigned to a condition in which they would have to make judgments for negative behaviors performed by Black people.

After the instructions and six practice trials with feedback, participants began the training phase, which consisted of 480 trials divided into six blocks of 80 trials. Appendix A contains an example of a typical trial in this condition. For each trial, the photograph of a Black young male was presented on the top of the screen, accompanied by the label "African American" to the left of the photograph, to insure that participants categorized the person by race. Underneath the photograph, participants saw a behavior that was strongly suggestive of a negative stereotypic trait as indicated by a pretest (loud, criminal, unintelligent, unreliable, irresponsible violent, dishonest, dangerous, lazy, promiscuous). After a 3,000ms exposure time to the photo and behavior, a situational and a dispositional explanation appeared on the left and right bottom of the screen. For example, the behavior "Arrived at work an hour late" was accompanied by a situational explanation ("The power went out and reset his alarm") and a dispositional explanation ("He is a particularly irresponsible person"). The participants' task

was to choose the situational explanation of the two, by pressing one of two designated keys on the keyboard. For half of the trials the situational explanation appeared on the left side of the screen, and for the other half on the right side of the screen.

Participants randomly assigned to the Grammar Control condition also completed 480 trials in which they were exposed to the same photographs and stereotypic African American behaviors as participants in the experimental condition. However, instead of making situational attributions for those behaviors, they were asked to count the number of nouns (240 trials) and verbs (240 trials) in the behavioral sentences and make two-choice decisions using the keyboard (for example, they had to choose between “2 or under 2 nouns” and ”over 2 nouns”). Appendix A contains an example of a typical trial in this condition.

Phase 3: Measure of situational inference activation. Following the training phase, all participants completed an adaptation of the probe recognition task, designed to measure spontaneous situational inference activation (Ham & Vonk, 2003).

Probe recognition task pretest. The goal of the pretest for this study was twofold: to generate behavioral sentences that allow for a relatively high level of activation of situational inferences, while also allowing for the activation of negative African American stereotypic traits not seen in training: agitated, bitter, uneducated, ignorant, impractical, poor, suspicious, superstitious. Eight research assistants generated several behaviors that reflected each of these traits. Of those behaviors, I chose the most appropriate ones to be pretested in terms of trait activation. Twenty-two undergraduate students participated in the pretest, as one means to fulfill a course requirement. They were presented with a total of 79 behavioral sentences and instructed to write down one-word inferences about the situation in which each

behavior occurred. After aggregating their responses, I selected 20 sentences that allowed for the strongest levels of activation of situational inferences. At the same time, attention was given that these sentences also allowed for a relatively high level of activation of negative African American trait inferences, as suggested by data from 18 participants who wrote down one-word trait inferences for each behavior. Table 2 presents the list of the chosen behaviors, situations and traits, and percentage of participants that generated each situation and trait (or close synonyms) after reading the sentence.

Table 2. Behaviors Used in the Probe Recognition Task (Study 1), Associated Situations and Traits, and Percentage of Pretest Participants Who Generated Each Situation (N = 22) and Trait (N = 18)

Behavior	Situation	Pretest Percentage	Trait	Pretest Percentage
Guards his bag when anyone passes him on the street	Robber	25%	Suspicious	41.18%
Installed a security camera in his front yard	Crime	25%	Suspicious	52.94%
Can not find a good job	Recession	25%	Uneducated	
Is not familiar with American history	Immigrant	50%	Uneducated	23.53%
Worked as a janitor, as he could not get hired anywhere else*	Economy	15%	Uneducated	11.76%
Wears tattered old jeans most of the time	Fashion	30%	Poor	25.53%
Did not have cash in his wallet at lunch	Credit	20%	Poor	29.41%
Walks to the supermarket for grocery shopping	Close	25%	Poor	

Does not own a car, instead he rides public transportation*	City	20%	Poor	17.65%
Could not unlock the door although he had the right key	Stuck	40%	Impractical	17.65% (Dumb)
Took the long way home from his sister's house	Traffic	10%	Impractical	
Could not get the computer started in his office	Broken	42.4%	Impractical	45.7%
Does not understand an important document	Unclear	20%	Ignorant	35.29% (Dumb)
Got an F on his final exam	Hard	45%	Ignorant	52.94% (Lazy)
Did not know who the vice-presidential candidates were for the current election*	Foreign	10%	Ignorant	29.41%
Falsely recited current law to a friend*	Changed	20%	Ignorant	23.53%
Didn't know how to start a lawnmower	New	20%	Ignorant	11.76%
Was pacing through his parents' living room*	Searching	15%	Agitated	47.06%
Threw a glass of water on Bill	Fire	25%	Bitter	23.53% (Rude)
Declined to get coffee with an old accomplished classmate.	Time	10%	Bitter	35.29% (Jealous)

* Behavior also used in Study 2

Probe recognition task. The task was modeled after Ham and Vonk (2003; Study 1) who adapted the probe recognition paradigm to measure the activation of spontaneous

situational inferences. Participants were told that they would complete a memory task, unrelated to the previous training task. After six practice trials, participants completed 120 trials in which they were presented, one at a time, with 20 behaviors that were suggestive of negative stereotypic African American traits not seen in training (suspicious, uneducated, poor, impractical, ignorant, agitated, bitter; see Table 2 for behaviors, situational probes and underlying traits). Each behavioral sentence appeared six times followed by one different probe type each time. The behavior/probe combinations appeared in random order on the screen. Two of the probe types were of interest to the current analyses. First, the experimental probes were situational probes that did not appear in the sentences but that were presumably spontaneously activated by the sentences, according to the pretest. Second, the control probes were situational probes that did not appear in the sentence and were not spontaneously activated by the sentences. Consistent with Ham and Vonk, these probes were situational probes activated by a different behavior in the task; however, they were rearranged to follow a behavior that did not imply them. The correct answer for both experimental and control probes was NO.

I also included four filler probes for each sentence. First, there were two probes that were actually seen in the sentence. The correct answer for these probes was YES. These probes were included so that participants would not learn that a successful responding strategy is to answer NO on every trial. Additionally, in order to keep participants focused on the meaning of the sentences, I included two additional filler probes that were verbs that were seen and not seen in the sentences. Table 3 contains an example of a behavior and associated probes

Table 3. Example of a behavior and associated probes used in Study 1

Probe	Correct Answer	Probe Type
		Got an F on his final exam.
Situational	No	Situation not seen but implied: *"Hard"*
Control	No	Situation not seen or implied: *"Broken"*
Filler 1	Yes	Property seen: *"Final"*
Filler 2	Yes	Property seen: *"Exam"*
Filler 3	Yes	Verb seen: *"Got"*
Filler 4	No	Verb not seen: *"Invited"*

Each trial started with a 1,000 ms exposure to a row of five X's in the middle of the screen, in order to focus participants' gaze. The fixation point was then replaced with a photo of a Black man not seen in training, paired with a behavioral sentence; both remained on the screen for 3,000 ms. A blank screen then appeared for 500ms, which was followed by the presentation of a probe in the middle of the screen. The YES and NO response options were displayed on the bottom of the screen. The probe remained on the screen until participants made a decision about whether the probe had appeared in the sentence or not, using the appropriate keys on the keyboard. After a response, a new trial began with a row of X's. Participants' responses were recorded in milliseconds.

Results

Preliminary analyses. I eliminated data from five participants for whom there were computer problems that interfered with the tasks. This resulted in a working N of 76 participants.

Response time data were prepared in accordance with Ham and Vonk (2003; Study 1). First, only correct responses were considered in the final analyses of response time data. Second, outliers below 200 ms and above 2,000 ms were dropped. Finally, response times were log-transformed to avoid a skewed distribution of response times. Two variables were obtained: the mean response time to situational probes (words that were not present but implied by the sentence) and the mean response time to control probes (words that were not present or implied by the sentence). Results are reported in milliseconds from the untransformed variables.

For accuracy data, I computed two variables: the total number of trials in which participants correctly rejected the situational probes and the total number of trials in which participants correctly rejected the control probes. As did Ham and Vonk (2003), I used a square root transformation to avoid skewed data. However, the pattern of findings did not differ between the transformed and untransformed analyses, so untransformed data are reported.

Main analyses.

Response times. I conducted a mixed-design 2 (Condition: Situational Attribution Training versus Grammar Control) X 2 (Probe: Situational versus Control) ANOVA, with the second factor as a repeated measure on the response times to situational and control probes. There was no main effect of condition, $F(1, 74) = .23$, $p = .64$, $\eta^2 = .003$, such that participants in the training and control conditions responded equally quickly. This analysis

revealed a significant main effect of probe, $F(1, 74) = 32.31$, $p < .001$, $\eta^2 = .30$, such that participants were significantly slower to reject the situational probes (M = 958.88) compared to the control probes (M = 932.03). Contrary to our hypothesis, this difference was not qualified by an interaction with the training condition, $F(1, 74) = .81$, $p = .37$, $\eta^2 = 01$. Participants were slower to reject the situational probes compared to the control probes both after completing Situational Attribution Training, $F(1, 42) = 12.97$, $p < .01$ (M = 988.62, SD = 412.04 and M = 945.11, SD = 315.63, respectively), and after completing the Grammar Control condition, $F(1, 32) = 9.56$, $p < .001$, $\eta^2 = .38$ (M= 952.92, SD = 225.05 and M = 880.04, SD = 215.34, respectively). In other words, both training and control participants spontaneously activated situational inferences for negative stereotypic African American behaviors.

Accuracy data. I conducted a mixed-design 2 (Condition: Situational Attribution Training versus Grammar Control) X 2 (Probe: Situational versus Control) ANOVA on accuracy data, with the second factor as a repeated measure. This analysis revealed no significant main effects or interactions, all *F*s < 1. Thus, I was not able to observe evidence of SSI activation in terms of accuracy in either the Situational Attribution Training condition, $F(1, 42) = 0$, $p = 1$, $\eta^2 = .00$ (M = 19.84 and M = 19.84, respectively) or the Grammar Control condition, $F(1, 32) = 1.85$, $p = .18$, $\eta^2 = .05$ (M = 19.85 and M = 19.94, respectively).

Discussion

Study 1 findings did not support my first hypothesis that Situational Attribution Training would increase the tendency to make SSIs. In the present experiment, training had no effect on SSIs above and beyond that of controls. Response time analyses revealed that participants who completed Situational Attribution Training spontaneously activated situational inferences, as

evidenced by quicker response times to rejecting the situational compared to the control probe. However, this tendency was also observed for participants in the control condition. Both effects were of medium sizes, suggesting similar levels of SSI activation for both training and control participants.

There are several theoretical and methodological implications of the null findings in the current study. This discussion will include an analysis of these issues, as well as suggested future directions for the investigation of situational inferences following Situational Attribution Training.

Given that SSIs did not differ between training and control participants, it may be fruitful to have a repeated-measures design, in which SSIs are assessed before and after participants complete Situational Attribution Training. This way it would be possible to assess whether SSIs have increased following training and not relative to control participants whose SSIs may have been activated during the completion of Grammar Control condition. Similar to the Situational Attribution Training condition, participants in the Grammar Control condition were also exposed to negative stereotypic behaviors performed by Black men. As suggested by findings from Ham and Vonk (2006) exposure to behavioral statements can determine the automatic activation of SSIs. An alternative option for future research would be to compare the SSI activation of training participants to that of control participants who only complete the dependent measure of inference activation without receiving any type of training. This would enable us to investigate the relative activation of SSIs for training compared to “pure” control participants.

Second, I was not able to see any training effects on accuracy rates, suggesting, once more, that training had no effect on SSIs. Some methodological issues may also be raised. In the current studies accuracy reached a ceiling effect (error rates below 1%), consistent with probe recognition task studies (Uleman et al., 1996a; Wigboldus et al.. 2003) that did not find significant effects on error rates. One exception comes from Stewart et al. (2003) who found significant effects on accuracy rates using a more challenging task in which participants were tested for their memory of three behavioral sentences in each trial. However, the current task may not be challenging enough for participants to make a significant number of errors. One solution to increase error rates would be to restrict the window of responding to the probe recognition task. This change would force participants to respond more quickly and possibly increase the chance of incorrectly identifying a trait that did not appear in the sentence.

Third, from a theoretical point of view, it may be that Situational Attribution Training does not have an effect on spontaneous situational inferences but only on intentional inferences. This explanation brings back the distinction between spontaneous inference research which studies automatic inferences (Gilbert & Malone, 1995; Krull & Erikson, 1995) and social inference research which studies intentional inferences (see Uleman et al., 1996b for a review). In the intentional social inference literature, Gilbert and Malone as well as Krull and Erikson proposed that people initially draw *trait* inferences when they are interested in understanding people, which is the case most of the time in our daily lives. Their models predict that when people have this goal in mind, they efficiently make trait inferences, which may be corrected later to take into account situational factors. However, Krull and Erikson also propose that people initially draw *situational* inferences when they are interested in understanding situations. Thus, the extent to which individuals make intentional trait or situational inferences depends

largely on their processing goals. This may not the case for spontaneous inferences such as STIs and SSIs, which are activated automatically, regardless of the perceiver's processing goals.

Based on this theoretical account, Ham and Vonk (1996) argued that individuals might activate both STIs and SSIs regardless of whether they want to understand the person or the situation. This proposition was supported by their 1996 findings, which showed that participants activated both STIs and SSIs for the same behaviors. Despite numerous findings that suggest that individuals underestimate situational explanations when making intentional inferences (see an overview of the correspondence bias research, Gilbert & Malone, 1995), participants in Ham and Vonk's studies showed SSI activation without any intervention that would stimulate this inference strategy. Consistent with these findings, control participants in the current study also showed reliable evidence of SSI activation in the absence of training. However, whereas training may not have a strong effect on SSIs, it may be possible that it has a significant effect on intentional situational inferences – situational attributions people make when they have an impression formation goal in mind. Specifically, it may be that by training participants to consider situational factors during Situational Attribution Training, we are changing participants' inference goals, such that they are more invested in understanding situations rather than traits. As a result, consistent with Krull and Erikson's model, they may strengthen their intentional situational inferences, which would negatively affect their tendency to draw quick, efficient trait inferences for negative stereotypic behaviors of Black men.

Finally, there was a great deal of variability in participants' response time data after completing Situational Attribution Training. This finding suggests that individual differences are an important aspect which should be taken into account when investigating the effects of training on implicit bias. I designed Study 2 to investigate this hypothesis.

Study 2

Stewart and colleagues (2010) showed that Situational Attribution Training reduced the degree of association between negative traits and Black photos. Using a different paradigm, in the current study I sought to replicate these findings by showing that training reduces the likelihood to make spontaneous trait inferences for negative stereotype-consistent behaviors performed by Black men. In addition, in the current study I looked at whether the effects of training have become automatic, such that the stereotyping reduction effects of training would be maintained even under conditions of cognitive load. Finally, taking a novel approach for the stereotype reduction literature, I investigated for whom training was most effective, by studying the moderators of Situational Attribution Training. This investigation of individual differences is made particularly important by the null findings of Study 1, and the suggestion that perhaps training will be effective for some but not all participants.

Participants and Design

White American students (N = 129) enrolled in Introductory Psychology courses participated in the first part of experiment, as one means to fulfill a course requirement. Of those, 117 participants returned to the lab for the second part of the experiment, which took place two to four days after the first part. In the first part of the experiment, all participants completed several questionnaires on the computer. In the second part, participants returned to the lab and were randomly assigned to one of four conditions: 2 Training groups (Situational Attribution Training Condition versus Grammar Control Condition) X 2 Load for Probe Recognition Task (High Load versus No Load).

Procedure and Materials

Phase 1: Moderators. In the initial phase of the experiment, participants completed several questionnaires that measured individual difference variables that were hypothesized to moderate the effects of Situational Attribution Training on decreased automatic stereotype activation. These measures included explicit prejudice, motivations to control prejudice, NFC, and PNS. Appendix B contains all the scales and instructions used in this study.

The Social Distance Scale (SDS, adapted from Bogardus, 1925) has been successfully used in the past to measure prejudice against Black people (Stewart et al., 2003). The scale consists of several statements such as, "I would be willing to have a Black American person as my roommate/friend/dance partner/governor/president, etc." The SDS contains fourteen such items, ranging from a very intimate relationship (e.g., spouse) to a more distant relationship (e.g. president). Participants rated their agreement with each of the items on a 7-point scale (1 – *Strongly disagree* to 7 – *Strongly agree*). Scores were added to obtain a final SDS score, *Cronbach* α = .92, such that lower scores indicate more prejudice.

The Attitude Towards Blacks Scale (ATBS; Brigham, 1993) was also used as a measure of prejudice. The ATBS is a 20-item scale and sample items include "I favor open housing laws that allow more racial integration of neighborhoods" and "I get very upset when I hear a White make a prejudiced remark about a Black person". Participants rated their agreement with those items on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). After reversing the scores of appropriate items, responses were totaled to obtain a final ATBS score, *Cronbach* α = .85, with higher scores indicating a more positive attitude toward Blacks.

The Modern Racism Scale (MRS; McConahay, 1983), a 5-item self-report questionnaire, was designed to measure the denial of current discrimination of Black people and lack of support for Black people's fight for equality (e.g., "Discrimination against Blacks is no longer a problem in the United States," "Blacks are too demanding in their push for equal rights"). Responses were measured on scales from 1 (*strongly disagree*) to 7 (*strongly agree*). Scores were added to obtain an overall MRS score, *Cronbach* α = .78, with higher values representing more modern racism.

The Internal / External Motivation to Control Prejudice Scale (IMS/EMS; Plant & Devine, 1998) is a self-report questionnaire that contains two subscales. Five items measure participants' internal motivation to respond without racial prejudice (e.g., "Being non-prejudiced toward Black people is important to my self-concept."). The other five items measure participants' external motivations to respond without racial prejudice (e.g., "I try to hide any negative thoughts about Black people in order to avoid negative reactions from others"). All items were measured on a 7-point scale, from 1 (*Strongly disagree*) to 7 (*Strongly agree*), and responses were added together to obtain an IMS total score, *Cronbach* α = .82 and an EMS total score, *Cronbach* α = .87, with higher scores indicating higher levels of either motivation.

The Need for Cognition Scale (NFC; Cacioppo, Petty, & Kao, 1984) measures participants' tendency to engage in and enjoy effortful thinking (e.g., "I would prefer complex to simple problems," "Thinking is not my idea of fun"). Participants were asked to rate their agreement with 18 items on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). After reversing the scores of the appropriate items, responses were added to obtain a final NFC score, *Cronbach* α = .91, with higher scores reflecting higher NFC.

The Personal Need for Structure Scale (PNS; Neuberg & Newsom, 1993) measures two related concepts: the extent to which people desire to establish structure in their life (e.g., “I like to have a place for everything and everything in its place”) and the manner in which people respond to a lack of structure (e.g., “I don’t like situations that are uncertain”). Participants were asked to state their agreement with each of 11 statements on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Scores were added to obtain a final PNS score, *Cronbach* α = .84, with higher scores indicating higher need for structure.

In order to conceal the study’s goal of measuring racially biased attitudes, I also included several filler items, which measured individual’s religiosity and their tendency to ruminate. These scales were not analyzed for this study.

Phase 2: Situational Attribution Training. Two to four days after they completed the questionnaires, participants returned to the lab to complete the second part of the study, which was presented as being unrelated to the first part. The second phase of Study 2 was identical with Phase 1 of Study 1, with participants being randomly assigned to either the Situational Attribution Training or the Grammar Control condition.

Phase 3: Measure of stereotype activation. Following the training phase, all participants completed the probe recognition task, as a measure of stereotypic trait inference activation.

Probe recognition task pretest. The goal of the pretest was to generate behavioral sentences that allow for a relatively high level of activation of negative African American stereotypic traits not seen in training: agitated, bitter, uneducated, ignorant, impractical, poor, suspicious, superstitious. The same 79 behavioral sentences generated by eight research assistants based on the negative stereotypic traits were pretested on 18

undergraduate students, who participated as a means to fulfill a course requirement. Participants were instructed to write down one-word inferences about the trait of the person they read about. Afterwards, I aggregated their answers and selected twelve sentences that allowed for the strongest levels of activation of African-American trait inferences. Table 4 presents the list of the chosen behaviors, traits, and percentage of participants that generated that trait (or close synonyms) after reading the sentence.

Table 4. Behaviors Used in the Probe Recognition Task (Study 2), Associated Traits, and Percentage of Pretest Participants Who Generated Each Trait (N = 18)

Behavior	Trait	Pretest Percentage
Was asked to name seven continents and he named seven countries	Uneducated	11.76%
Worked as a janitor, as he could not get hired anywhere else	Uneducated	11.76%
Could not pay for his lunch today	Poor	55.82%
Does not own a car, instead he rides public transportation	Poor	17.65
Was pacing through his parents' living room	Agitated	47.06%
Shakes his leg continuously while sitting at the table	Agitated	64.70%
Did not know who the vice-presidential candidates were for the current election	Ignorant	29.41%
Falsely recited current law to a friend	Ignorant	23.53%
Hasn't congratulated a coworker who received a promotion over him	Bitter	17.65%
Resents the fact that his father remarried	Bitter	23.53%
Did not come to work on Friday the 13th	Superstitious	70.58%
Walks around the ladder on his way home	Superstitious	58.82%

Probe recognition task. The task was modeled after Wigboldus et al. (2003; Study 2), and it was introduced to participants as a memory task, unrelated to the previous task. After two practice trials, participants were presented, one at a time, with 12 behaviors that were

suggestive of negative stereotypic African American traits not seen in training (uneducated, poor, agitated, ignorant, bitter, superstitious). Each behavior was presented twice, once paired with a Black photo not seen in training, and once paired with a White photo. The presentation order for all behaviors was randomized. The experimenter informed participants that their memory would be tested after the presentation of each sentence. After a 1,200 ms exposure time to the behavioral sentence and the photo, 5 one-word probes were presented one by one. The participants' task was to decide as quickly as possible whether the probe had literally appeared in the sentence. There were two probes of interest: a trait probe that was not present, but that was presumably activated while reading the sentence (e.g., "uneducated" for "Worked as a janitor, as he could not get hired anywhere else") and a control probe, which was a non-stereotypic trait that was not present or activated by reading the sentence (e.g., "deceptive"). Other filler probes contained a verb that was not present in the sentence, a verb that was present in the sentence, and an article or preposition that was present in half of the sentences and not present in the rest. Thus, for half of the trials the correct answer was "yes," and for the other half the correct answer was "no." The order of the probes for each sentence was fully randomized. Table 5 presents an example of a behavior and associated probes.

Table 5. Example of a behavior and associated probes used in Study 2

Probe	Correct Answer	Probe Type
		Could not pay for his lunch today.
Trait	No	Stereotypic trait not seen but implied: *"Poor"*
Control	No	Nonstereotypic trait not seen or implied: *"Nervous"*
Filler 1	Yes	Verb seen: *"Pay"*
Filler 2	Yes	Property seen: *"Lunch"*
Filler 3	Yes	Preposition seen/not seen: *"For"/"Of"*

Load manipulation. The load manipulation was modeled after Wigboldus et al. (2003) who successfully used it in conjunction with the probe recognition task[1]. Half of the participants completed the probe recognition task under a no-load condition. The other half completed the task under conditions of high cognitive load. Before each trial (behavior and five probes), participants were presented with a five-digit number for 5,000ms. At the end of each trial, they were asked to recall this number in writing. A new number was then presented before the onset of the following trial. Presumably, participants were rehearsing the number while performing the task, and were thus cognitively loaded.

Results

Preliminary analyses. Based on *a priori* criteria, such as low memory rates for load task, experimenter or task error, I eliminated several participants from the initial data set. Data from three participants who had extremely low memory rates for the load task (0%, 0%, and

4.17% correct recall) were initially eliminated. Also, 26 additional participants were eliminated due to experimenter error, falling asleep during the training phase, hitting the wrong keys and not completing the training task completely, and computer problems during the task. After the elimination, the final N was 88. Participants in the final sample had an age that ranged between 18 and 35 years, with a mean of 20 years. Two of the participants were non-native English speakers. Thirty participants were men and 58 were women.

The probe recognition task data were prepared in accordance with Wigboldus et al. (2004, Study 2), after which I modeled the task. Of interest were the response time data from the trait and control probes, to which the correct answer was “no.” First, I eliminated all incorrect “yes” responses. This resulted in the elimination of 1.59% of responses (67 out of 4224 total responses). Second, I systematically eliminated outliers, by replacing response times greater than two standard deviations above the mean with the mean for that particular item. Response times for relevant items were averaged to obtain two scores: the response time to trait probes and to control probes after being presented with a negative behavior paired with a Black photo. A difference score was also computed by subtracting the response time to the control probes from the trait probes; this variable was named the stereotyping score. If participants automatically activate negative stereotypes of Blacks, they should take longer to reject the negative trait probe compared to control probe. Thus, higher stereotyping scores reflect more stereotyping of African Americans, with negative stereotyping scores suggesting lower automatic stereotyping.

Main analyses.

Stereotype reduction as a function of training and cognitive load. I first conducted a mixed-design ANOVA, with the type of probe (trait vs. control) and race of photo (Black vs. White) as repeated factors, and training condition (Situational Attribution Training vs. Grammar Control) and Cognitive Load (no load vs. high load) as between-subjects variables. Results showed a main effect of load condition, $F(1, 84) = 17.29$, $p < .001$, $\eta^2 = .17$, such that loaded participants were significantly slower (M = 1003.45) compared to non-loaded participants (M = 837.94), across all trials and training conditions. Results also revealed a main effect of type of probe, $F(1, 84) = 7.13$, $p = .009$, $\eta^2 = .08$, such that participants were significantly slower to reject the trait probe (M = 933.85) compared to the control probe (M = 907.53), regardless of the race of the photograph paired with the behavior. This main effect was qualified by an interaction trending towards significance between type of probe and cognitive load condition, $F(1, 297) = 3.28$, $p = .07$, $\eta^2 = .04$. Follow up analyses showed that loaded participants were significantly slower to reject all trait probes (M = 1024.01) compared to all control probes (M = 980.29), $F(1, 41) = 7.59$, $p = .009$, $\eta^2 = .16$. However, this difference was not significant for the no load participants, $F(1, 45) = .53$, $p = .47$, $\eta^2 = .01$ (M = 842.17 and M = 833.70 for the trait and control probes, respectively). These findings suggest that our load manipulation was partially successful in cognitively loading participants. However, we observed no stereotyping effects, with participants not being slower to reject the trait compared to the control probe in either condition. In addition, there were no stereotyping reduction effects of training condition. However, these effects may be moderated by several individual differences variables, such as explicit prejudice, motivations to control prejudice, NFC, and PNS.

I initially investigated the role of individual differences in moderating the effects of training on stereotype activation by computing correlations between the stereotyping score, explicit prejudice, motivations to control prejudice, NFC, and PNS for each condition. Tables 6 and 7 display these correlations by training and load conditions. The relationship between all individual differences, training, and load condition were also investigated using hierarchical multiple regression analyses. Significant effects were only found for modern racism, NFC, and IMS/EMS. I reported these analyses below.

Table 6. Means, Standard Deviations, and Intercorrelations for Situational Attribution Training Participants who completed the probe recognition task under No Cognitive Load (Above the Diagonal) and Cognitive Load (Below the Diagonal) Conditions in Study 2, N = 47

	Non Loaded Participants (*N* = 23)		Loaded Participants (*N* = 24)		Intercorrelations							
Variables	M	SD	M	SD	1	2	3	4	5	6	7	8
1. Stereotyping Score	19.61	101.57	79.51	170.86	-	-.22	-.11	.42*	.36†	.23	-.39†	.30
2. Social Distance	89.00	14.67	91.92	8.19	.07	-	.81*	-.47*	.42*	-.34	.34	-.32
3. Attitude Towards Blacks	109.00	17.94	114.42	14.11	.06	.60*	-	-.69*	.56*	-.23	.42*	-.10
4. Modern Racism	11.74	4.98	10.21	5.00	.18	-.34	-.77*	-	-.27	.15	-.45*	-.03
5. IMS	28.26	7.11	29.75	4.81	.10	.28	.63*	-.54*	-	.25	-.08	-.14
6. EMS	21.74	8.15	18.42	7.90	-.06	-.41*	-.33	.49*	-.17	-	-.50*	.24
7. Need for Cognition	82.17	20.15	78.96	18.03	.03	.15	.34	-.28	.18	.03	-	-.35
8. Personal Need for Structure	42.52	12.49	41.04	9.81	.01	-.24	-.34	.16	-.08	-.01	-.46*	-

†p<.10, *p<.05

Table 7. Means, Standard Deviations, and Intercorrelations for Grammar Control Participants who completed the probe recognition task under No Cognitive Load (Above the Diagonal) and Cognitive Load (Below the Diagonal) Conditions in Study 2, N = 41

	Non Loaded Participants (*N* = 23)		Loaded Participants (*N* = 18)		Intercorrelations							
Variables	M	SD	M	SD	1	2	3	4	5	6	7	8
1. Stereotyping Score	38.32	89.24	33.75	125.74	-	.08	.09	-.28	.28	-.37†	.08	.15
2. Social Distance	91.52	11.36	90.00	9.00	.10	-	.72*	-.49*	.68*	-.09	.35	-.01
3. Attitude Towards Blacks	116.52	12.28	105.67	15.79	-.19	.76*	-	-.60*	.70*	-.29	.35	.01
4. Modern Racism	9.74	4.28	12.39	6.44	-.07	-.53*	-.72*	-	-.59*	.35	-.24	.04
5. IMS	29.65	5.66	28.17	5.77	.19	.34	.52*	-.27	-	-.37	.19	-.02
6. EMS	18.65	5.92	20.11	7.80	.38	.23	-.05	-.10	.26	-	-.06	-.04
7. Need for Cognition	91.48	13.90	82.83	15.63	-.36	.28	.28	-.02	.07	.31	-	.32
8. Personal Need for Structure	43.39	7.85	42.55	10.98	-.31	-.03	.08	-.03	.19	.07	.05	-

†p<.10, *p<.05

Need for cognition and stereotype reduction. Because the dependent variable in this analysis was a repeated measure variable – the response times to control probes and trait probes - I used the Sum/Difference regression model (Judd, Kenny, & McClelland, 2001), which allowed for the examination of interactions with the repeated-measure variable. Consistent with the Sum/Difference regression model, I conducted two regression analyses with two different dependent measures. In both models I conducted hierarchical multiple regression analyses with NFC (mean centered), training condition (dummy coded: control = 0; training = 1), and load condition (dummy coded: no load = 0, load = 1) as predictors in the first step. In the second step I added three two-way interaction terms between NFC, training condition and load condition. A third step added the three-way interaction between NFC, training, and load conditions. The two models differed however in the dependent variable used. In the between-subjects model, the dependent variable was the sum between the response times to trait probes and control probes. This model enabled me to examine the main effects and interactions of NFC, training, and load on the overall reaction time to both trait and control probes. Table 8 presents standardized and unstandardized regression coefficients, as well as standard errors for main effects and interaction terms in all three regression steps for the between-subjects model. Analyses revealed a main effect of load condition, such that loaded participants were significantly slower to respond to all probes (trait and control) compared to those who were not loaded. No other main effects or interactions were significant.

Table 8. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Need for Cognition (NFC), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3	
Variable	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
NFC	-1.44 (2.63)	-.06	1.24 (5.32)	.05	-.23 (6.51)	-.01
Training	-17.14 (90.98)	-.02	46.31 (131.88)	.05	36.32 (134.97)	.04
Load	299.25 (90.39)	.34*	-369.88 (139.94)	.42*	360.26 (142.77)	.41*
NFC X Training			-2.74 (5.67)	-.09	-.58 (7.91)	-.02
NFC X Load			-1.63 (5.38)	-.04	1.34 (9.27)	.04
Training X Load			121.00 (189.46)	-.12	-116.65(190.77)	-.12
NFC X Training X Load					-4.50 (11.41)	.10

*$p < .05$

Next, I investigated the within-subjects model, with the difference score between response times to trait and control probes as a dependent variable. As a reminder, being slower at responding to trait versus control probes in the probe recognition task denotes negative stereotyping trait activation. Thus, higher difference scores between the response times to trait versus control probes represent greater stereotyping of African Americans. Overall, participants did not show evidence of implicit stereotyping, as there was not a significant difference in their response time to trait versus control probes, as suggested by the statistics of the constant, b = 28.96, SE = 23.60, p = .22. Table 9 presents standardized and unstandardized regression coefficients, as well as standard errors for main effects and interaction terms in all three

regression steps for the within-subjects model. Findings from the third regression step showed an interaction trending towards significance between NFC, training condition, and load condition, with the addition of this three-way interaction in the third regression step adding incremental variance that was approaching significance.[2] To break down this three-way interaction, I analyzed the simple slopes for the relationship between NFC and stereotyping for each condition. Figure 1 is a visual representation of this interaction. For the Training/No Load participants, the simple slope was negative, $b = -1.95$, $SE = 1.33$, $p = .15$, such that higher NFC was associated with lower stereotyping after Situational Attribution Training. Control/No load participants showed a different, less dramatic pattern, $b = .50$, $SE = 1.93$, $p = .80$, as did Training/Load participants, $b = .33$, $SE = 1.46$, $p = .82$. Surprisingly, for the Control/Load participants, higher NFC was also associated with lower stereotyping, $b = -2.94$, $SE = 1.95$, $p = .14$. Overall, findings suggest that Situational Attribution Training worked best for participants high in NFC who were not cognitively loaded, as they showed reduced automatic stereotype activation. Thus, it seems that the effects of training are maximized for participants who are high in NFC and have enough cognitive resources.

Table 9. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Need for Cognition (NFC), Training Condition, and Load Condition Predicting Stereotyping Scores, the Difference Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3	
Variable	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
NFC	-.93 (.80)	-.13	-1.36 (1.60)	-.19	.50 (1.93)	.07
Training	5.49 (27.63)	.02	-30.89 (39.77)	-.12	-18.20 (40.03)	-.07
Load	24.80 (27.45)	.10	-16.01 (42.20)	-.06	-3.79 (42.35)	-.01
NFC X Training			.30 (1.71)	.03	-2.45 (2.35)	-.27
NFC X Load			.33 (1.62)	.03	-3.44 (2.75)	-.32
Training X Load			74.14 (57.14)	.26	68.62 (56.59)	.24
NFC X Training X Load					5.72 (3.38)	.43*

Note. ΔR^2 = .02 for Step 2 (p =.64) and ΔR^2 = .03 for Step 3 (p = .09).

*p = .09

Modern racism and stereotype reduction. Similar to the NFC analyses, I used the Sum/Difference regression model to investigate the role of modern racism, training, and load condition on the response times to trait relative to control probes. In the between-subjects model I entered the sum between the response times to trait and control probes as a dependent measure and modern racism, training condition (dummy coded; control = 0, training = 1), and load condition (dummy coded; no load = 0, load = 1) as predictors in the first regression step. A second step included the three two-way interaction terms between the predictors, and the third

step included the three-way interaction term between modern racism, training condition, and load condition. Table 10 presents standardized and unstandardized regression coefficients, as well as standard errors for main effects and interaction terms in all three regression steps for this model. Similar to the between-subjects model for the NFC analyses, results revealed a main effect of load, such that participants who were cognitively loaded while completing the probe recognition task were significantly slower to respond to both trait and control probes, compared to participants who were not cognitively loaded. No other main effects or interactions were significant.

Table 10. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Modern Racism (MRS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3	
Variable	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
MRS	3.53 (8.64)	.04	-7.48 (17.15)	-.09	-19.74 (21.00)	-.23
Training	-7.49 (89.30)	-.01	39.84 (127.36)	.05	47.18 (127.54)	.05
Load	306.10 (89.26)	.35*	377.87 (136.82)	.43*	382.32 (136.87)	.44*
MRS X Training			18.00 (17.98)	.15	39.29 (27.67)	.33
MRS X Load			1.91 (18.18)	.02	21.16 (26.31)	.19
Training X Load			-98.70 (184.96)	-.10	-102.05 (184.97)	-.10
MRS X Training X Load					-36.84 (36.40)	-.22

Note. ΔR^2 = .06 for Step 2 (p =.14) and ΔR^2 = .004 for Step 3 (p = .54).

*p = .05

Next, I investigated the within-subjects model, within a hierarchical multiple regression analysis, which contained the same predictors as the between-subjects model. However, this model used the difference score between the response times to trait and control probes as a dependent measure. Similar to the NFC analyses, there was no evidence of implicit stereotyping across conditions: participants were equally fast at responding to trait versus control probes, as suggested by the constant in the first regression step, b = 23.40, SE = 23.20, p = .32. Table 11 presents standardized and unstandardized regression coefficients, as well as standard errors for

main effects and interaction terms in all three regression steps for the within-subjects model. Findings from the second regression step showed a significant interaction between modern racism and training condition[2]. This interaction was trending towards significance in the third regression step (p = .09). Follow-up analyses showed that the slope of modern racism for training participants was approaching significance, *b*= 6.25, *SE* = 3.71, *p* = .09, such that higher modern racism was associated with more stereotyping after completing Situational Attribution Training. For control participants, the slope of modern racism was not significant, *b* = -2.97, *SE* = 3.66, *p* = .42, such that the level of modern racism was not significantly related to implicit stereotyping for participants who did not undergo Situational Attribution Training. Figure 2 presents a graphic representation of this interaction at high and low levels of modern racism (one standard deviation below and above the mean). Overall, findings suggest that the stereotyping reduction effects of Situational Attribution Training were most pronounced for participants who were initially low in modern racism, regardless of whether they were cognitively loaded or not. Importantly, training seemed to increase implicit bias for individuals who were high in modern racism.

Table 11. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Modern Racism (MRS), Training Condition, and Load Condition Predicting the Difference Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3	
Variable	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
MRS	1.47 (2.64)	.06	-3.60 (5.10)	-.15	-5.85 (6.27)	-.24
Training	11.69 (27.26)	.05	-19.99 (37.87)	-.08	-18.65 (38.07)	-.07
Load	29.51 (27.25)	.12	3.71 (40.68)	.01	4.52 (40.85)	.02
MRS X Training			10.52 (5.35)	.30*	14.42 (8.26)	.41
MRS X Load			.88 (5.41)	.03	4.40 (7.85)	.14
Training X Load			67.42 (54.99)	.24	66.81 (55.21)	.24
MRS X Training X Load					-6.75 (10.86)	-.14

Note. ΔR^2 = .06 for Step 2 (p =.14) and ΔR^2 = .004 for Step 3 (p = .54).

*p = .05

Motivations to respond without prejudice and stereotype reduction. Similar Sum/Difference analyses were conducted to investigate the effects of IMS and EMS on responses to Situational Attribution Training. In the between-subjects model I entered the sum between the response times to trait and control probes as a dependent measure and IMS, EMS, training condition (dummy coded; control = 0, training = 1), and load condition (dummy coded; no load = 0, load = 1) as predictors in the first regression step. A second step included the five two-way interaction terms between the predictors, and the third step included three-way interaction terms (IMS, EMS, training condition; IMS, training condition, load condition; EMS,

training condition, load condition). In the fourth step I added the interaction term between all predictors. Table 12 presents standardized and unstandardized regression coefficients, as well as standard errors for main effects and interaction terms in all four regression steps for this model. Consistent with previous analyses, there was a main effect of load, such that participants who were cognitively loaded while completing the probe recognition task were significantly slower to respond to both trait and control probes, compared to participants who were not cognitively loaded. In addition, there was a main effect of EMS, such that higher EMS was associated with slower responding to both trait and control probes.

Table 12. Summary of Hierarchical Regression Analyses for the Between-Subjects Model, with Internal Motivation to Control Prejudice (IMS), External Motivation to Control Prejudice (EMS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3		Step 4	
Variable	*B (SE B)*		*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
IMS	-4.47 (7.45)	-.06	-11.99 (13.99)	-.16	-13.09 (16.63)	-.17	-13.06 (16.55)	-.17
EMS	-13.59 (5.84)	-.23*	-7.03 (12.21)	-.12	-25.09 (17.09)	-.43	-25.08 (17.01)	-.43
Training	3.93 (87.08)	.01	19.75 (90.41)	.02	17.83 (90.45)	.02	-1.75 (91.22)	-.01
Load	292.92 (87.08)	.34*	262.58 (89.88)	.30*	244.47 (93.48)	.28*	248.57 (93.07)	.29*
IMS X Training			13.09 (16.10)	.13	9.30 (20.91)	.09	9.17 (20.81)	.09
IMS X Load			-.64 (16.33)	-.01	-4.41 (25.09)	-.04	-4.42 (24.97)	-.04
EMS X Training			-18.12 (12.92)	-.24	5.77 (20.81)	.08	10.05 (20.96)	.13
EMS X Load			8.15 (12.64)	.10	36.46 (21.52)	.45	36.42 (21.41)	.45
IMS X EMS			-1.03 (1.01)	-.11	1.01 (2.02)	.10	.99 (2.01)	.10
Training X IMS X EMS					-2.33 (2.45)	-.20	-1.18 (2.59)	-.10

Training X Load X IMS	14.07 (33.40)	.08	17.88 (33.36)	.10
Training X Load X EMS			-39.17 (27.28)	-.37
Training X Load X IMS X EMS			-3.82 (2.89)	-.17

Note. ΔR^2 = .03 for Step 2 (p =.63) , ΔR^2 = ..03 for Step 3 (p = .35), and ΔR^2 = .02 for Step 4 (p = .19).

*p < .05

In the within-subjects model I entered the same predictors as in the between-subjects model; the dependent measure was the difference score between the response times to control and trait probes. Table 13 presents standardized and unstandardized regression coefficients, as well as standard errors for main effects and interaction terms in all three regression steps for the within-subjects model. Results revealed a significant main effect of IMS, such that higher IMS was associated with more stereotyping. Findings from the second regression step also revealed a significant interaction between IMS and EMS. To follow up on the interaction between IMS and EMS, following recommendations from Aiken and West (1991), I investigated the slopes of IMS at low and high levels of EMS, calculated at one standard deviation below and above the mean of EMS, respectively. The simple slope of IMS at high levels of EMS was not significant, $b = -2.21$, $SE = 3.26$, $p = .50$. However, the simple slope of IMS at low levels of EMS was significant, $b = 10.15$, $SE = 2.87$, $p = .001$. Higher IMS was associated with more stereotyping of Black men, but only at low levels of EMS. Discrepant with the Devine and her colleagues (2002), individuals characterized by high IMS and low EMS were the most likely to show implicit bias towards Black targets, and this tendency was the same regardless of whether they completed Situational Attribution Training or not. No other main effects or interactions were significant.

Table 13. Summary of Hierarchical Regression Analyses for the Within-Subjects Model, with Internal Motivation to Control Prejudice (IMS), External Motivation to Control Prejudice (EMS), Training Condition, and Load Condition Predicting the Sum Between Response Times to Trait and Control Probes in Study 2 (N = 88)

	Step 1		Step 2		Step 3		Step 4	
Variable	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*	*B (SE B)*	*ß*
IMS	4.74 (2.30)	.22*	3.21 (4.13)	.15	3.07 (4.99)	.14	3.08 (4.94)	.14
EMS	.58 (1.80)	.03	-.14 (3.61)	-.01	-2.78 (5.13)	-.16	-2.78 (5.08)	-.16
Training	11.19 (26.82)	.12	22.51 (26.71)	.09	22.21 (27.16)	.09	15.08 (27.24)	.06
Load	30.21 (26.81)	.12	22.26 (26.55)	.10	22.63 (28.07)	.09	24.12 (27.80)	.10
IMS X Training			2.36 (4.76)	.08	1.76 (6.28)	.06	1.72 (6.22)	.06
IMS X Load			-1.69 (4.83)	-.05	-2.30 (7.54)	-.07	-2.31 (7.46)	-.07
EMS X Training			-2.76 (3.82)	-.13	.75 (6.25)	.03	2.31 (6.26)	.11
EMS X Load			4.43 (3.74)	.19	8.59 (6.46)	.36	8.58 (6.39)	.34
IMS X EMS			-.94 (.30)	-.34*	-.65 (.61)	-.23	-.65 (.60)	-.23
Training X IMS X EMS					-.34 (.74)	-.10	.08 (.77)	.02

Training X Load X IMS	2.17 (10.03)	.04	3.56 (9.96)	.07
Training X Load X EMS	-5.83 (8.23)	-.19	-5.69 (8.15)	-.18
Training X Load X IMS X EMS			-1.39 (.86)	-.22

Note. ΔR^2 = .11 for Step 2 (p =.07), ΔR^2 = .01 for Step 3 (p = .85), and ΔR^2 = .03 for Step 3 (p = .11).

*p = .05

Discussion

In the current study I addressed several questions about the mechanisms underlying Situational Attribution Training. This study did not replicate the previous studies which found reduced automatic stereotyping as a function of situational attribution training (Stewart et al., 2010). Overall, participants' stereotyping scores did not show a significant difference between training and control condition. Moreover, there was no evidence of automatic stereotype activation in either the training or the control condition.

Two additional findings are of particular importance and they refer to the moderating role of individual difference variables on training effects. First, findings suggest that the efficiency of Situational Attribution Training depends on both personal and situational constraints related to cognitive complexity. Moderation analyses revealed that training works best in reducing automatic stereotype activation when participants are high in NFC and are not cognitively loaded. Why do personal and situational cognitive complexity variables play such an important role? One explanation is that the Situational Attribution Training technique accomplishes its goal by teaching participants to take into account complex factors that may determine negative behaviors performed by Black men. This strategy poses a cognitive challenge for individuals, because they are required to process more information than in regular circumstances. Thus, individuals high in NFC, who are naturally inclined to prefer cognitive challenges, had the easiest time with this challenge and consequently showed the most benefits from training. Compared to low NFC people, individuals high in NFC may be more predisposed for the cognitively complex strategy of taking into account situational factors of out-group members' behaviors. The finding that these effects are only maintained under conditions of no

cognitive load suggests that the tendency to correct for stereotypic trait inferences has not become automatic even for these individuals.

One unexpected pattern of responding was observed for cognitively loaded participants in the control condition. For these participants, higher NFC was associated with less stereotyping. Why would higher NFC lead to less stereotyping for loaded controls? This effect is unclear; however, one explanation may come from previous research on cognitive load and stereotype activation. Contrary to the idea that cognitive load increases the tendency to rely on simple cognitive structures such as stereotypes, Gilbert and Hixon (1991) paradoxically found that loaded participants showed less stereotype activation compared to non-loaded participants. Their argument was that stereotype activation requires some cognitive or attentional resources. People who are depleted of those resources are less likely to show stereotype activation. Thus, it may be normal for our control participants who are loaded to show little stereotype activation. But why would high NFC individuals show this tendency more than low NFC? One explanation is that high NFC individuals enjoy processing information and they require cognitive resources to do so. When these resources are depleted, these high NFC individuals' stereotype activation suffers even more, such that they are especially less likely to show stereotype activation.

Overall, the findings related to the role of training, cognitive load, and NFC should be interpreted with caution because the overall interaction between these three variables only approached significance. Thus, these findings should be interpreted mainly as a trend and future studies should be conducted to replicate these effects using a different dependent measure. A discussion of future research ideas that address these limitations is included in the general discussion.

Second, Situational Attribution Training seemed to increase high in modern racism participants' stereotyping levels. Generally, individuals high in modern racism explicitly state that racial discrimination does not exist anymore and thus are unsupportive of Black's fights for equal rights. Not surprisingly, these individuals did not internalize the tendency to attribute negative behaviors of Black men to situational factors and continued to rely on traits when explaining negative behaviors of Black men. If anything, their high stereotyping scores suggest that training may elicit reactance. The exposure to negative stereotypic behaviors of African Americans coupled with the required task of choosing situational explanations may have had an irritating effect on these participants, ironically leading to an increase in automatic stereotype activation. This finding and interpretation are consistent with previous racial bias reduction research – Kawakami and colleagues (2005) found that a variation of the stereotype negation training led to reduced stereotype application as long as participants were not trying to avoid being influenced by training. The current study adds to our knowledge of reactance in stereotype reduction interventions by suggesting that high explicit prejudice increases the likelihood of such adverse reactions.

Overall, the current findings are consistent with research by Greenberg and Rosenfield, (1979) as well as Wittenbrink, et al., (1997) who showed that individuals high in explicit prejudice are more likely to engage in the ultimate attribution error. The current study suggests that high-prejudice participants are also more resistant to changing this attributional pattern, thus leading to less implicit bias reduction compared to low-prejudice participants.

The role of individual differences in moderating bias reduction intervention effects has been rarely documented in the literature (Levy, 1999). In the current research I addressed this understudied area of prejudice reduction research and showed that Situational Attribution

Training works best for individuals high in NFC and low in explicit prejudice. Despite their novelty and importance, these findings should be interpreted with caution because, overall, the current study was not able to fully replicate the Stewart and colleagues' (2010) findings. Specifically, in the current study I did not find a difference in stereotype activation between training and control participants. In fact, there was no evidence of stereotypic trait activation for either control or training participants. One reason for the discrepancy between the current research and the Stewart and colleagues findings may reside in the change of dependent measure. Unlike previous studies in which we used the person categorization task, in the current study I used the probe recognition task to measure changes in implicit racial bias. It may be that this measure is not sensitive enough to capture changes in stereotype activation following an intervention. In fact, there are no implicit bias reduction studies to date that used the probe recognition task as a dependent measure. The majority of studies from this literature employ quick categorization tasks such as sequential priming tasks (for example the person categorization task, Kawkami et al., 2000; Stewart et al., 2010), the IAT (Kawakami et al., 2000), or the weapons identification task (Stewart & Payne, 2008). Moreover, some studies using the probe recognition task have had difficulty in finding automatic stereotype activation for racial stereotypes even in the absence of any intervention. For example, Stewart et al. (2003) has found stereotyping effects with the probe recognition task only for positive, but not negative Black stereotypic traits.

Another indication that the dependent measure used in the current study may be problematic comes from the failure to replicate previous findings documenting the relationship between motivations to control prejudice and implicit racial bias. Devine and her colleagues (2002) found that individuals who are at the same time high in IMS and low in EMS display the

least implicit bias towards Blacks. Latu and colleagues (in press) replicated these findings for implicit gender bias. In contrast to these existing studies, the current research revealed that high IMS /low EMS participants showed the greatest level of implicit race bias, using the probe recognition task as a measure. This opposite pattern of findings suggests that our measure taps into different aspects of implicit racial bias than sequential priming tasks commonly used in implicit bias and implicit bias reduction research.

It is also important to speculate whether the failure to replicate Stewart et al. (2010) suggests that their findings are unreliable and thus Situational Attribution Training is not an effective tool in reducing automatic stereotype activation. However, it is not immediately apparent what the difference between the current studies and prior studies that have found support for the training's effectiveness (Stewart et al., 2010) might be, with the exception of the dependent measure used – the person categorization task in Stewart and colleagues and the probe recognition task in the current study. Otherwise, the studies had the same participant pool, used the same version of training, and were conducted within a two-year period by experimenters that were trained in a similar fashion. The finding that basic automatic stereotyping effects were not found for control participants in the current study (unlike for Stewart et al.'s controls), suggests that the probe recognition task is not appropriate for capturing the hypothesized effects and that Stewart et al. findings may be more reliable than the current findings. More research is needed, however, to address this issue. Future studies should investigate whether the Stewart et al. findings replicate using other measures of automatic stereotype activation, such as the lexical decision task or the IAT.

If the failure to replicate previous findings for training effectiveness and IMS/EMS is task-dependent, a different method for assessing implicit bias should be employed in future

research. The probe recognition task was employed in this study because of its successful history of being used in conjunction with a cognitive load task. A pretest revealed that the person categorization task, which was used in previous Situational Attribution Training studies, was not appropriate for the current goals. Theoretically, one hypothesis is that participants who are highly loaded should show more automatic stereotype activation compared to low-load participants, because depletion of cognitive resources should be associated with a greater tendency to rely on schemas and stereotypes. Gilbert and Hixon (1991) found the opposite pattern of results, with loaded participants showing more stereotype activation compared to non-loaded participants. Inconsistent with either of these two hypotheses, neither low nor high-load pretest participants showed automatic stereotype activation when using the person categorization task. This finding may not be surprising for the high-load participants – consistent with Gilbert and Hixon (1991), cognitive load sometimes has the paradoxical effect of decreasing stereotype activation. However, at a minimum, low-load participants should have shown automatic stereotype activation in order to proceed with the person categorization task as a dependent measure in the current studies. In addition, a significant difference between low and high load conditions on stereotype activation in the pretest would have motivated its use in conjunction with the Situational Attribution Training technique.

One problem may be that sequential priming tasks such as the person categorization task are designed to tap into automatic associations. As a consequence, adding cognitive load to assess the degree of automaticity of a process may lead to a ceiling effect of automaticity, thus confounding the results. Thus, one possibility for future research is to investigate the effects of training on intentional trait and situational inferences, by having participants make conscious attributions of negative stereotypic behaviors performed by Black men. As suggested in the

Study 1 discussion, this methodology would allow for the investigation of the effects of training on intentional inferences, and subsequently of processing goals. Half of the participants should complete this task under regular conditions and half under conditions of high cognitive load, to investigate whether the increase or decrease in the tendency to make situational or trait inferences is becoming more automatic following Situational Attribution Training. Adding cognitive load to an intentional inference task would allow us to investigate even slight increases in the automaticity of this process.

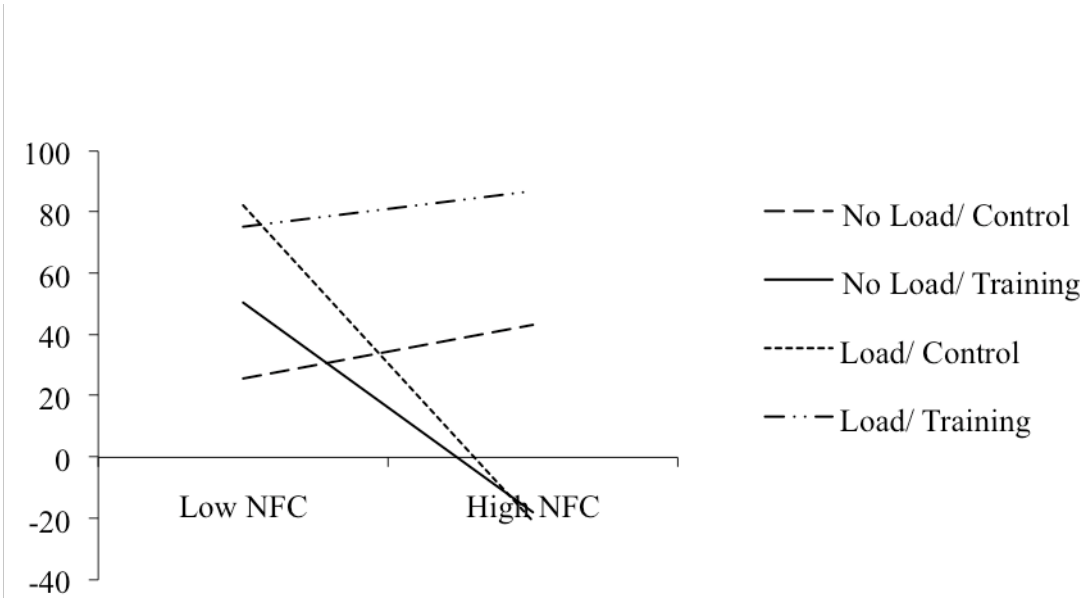


Figure 1. Stereotyping Level by Level of Need for Cognition (NFC), Training and Load Condition

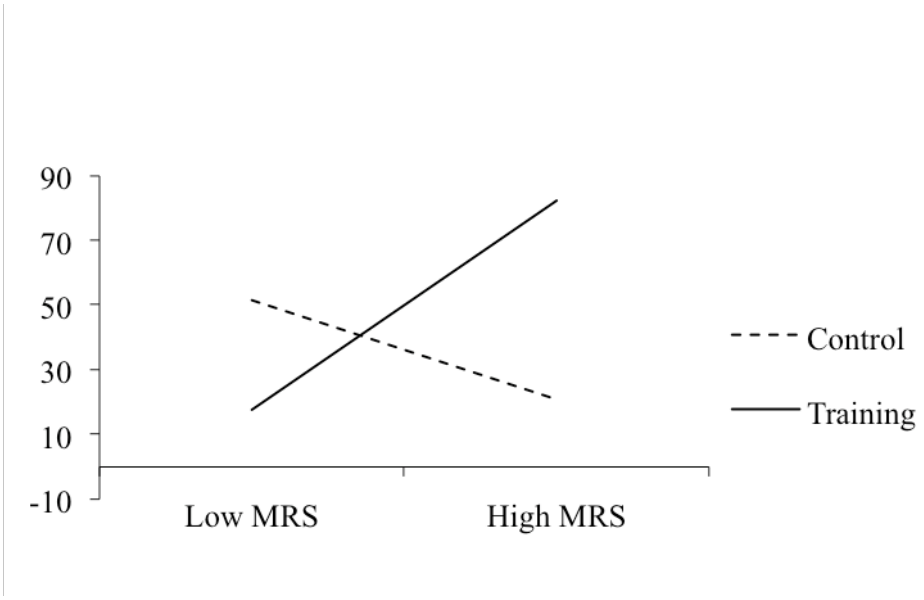


Figure 2. Stereotyping Level by Training Condition and Initial Level of Modern Racism

CHAPTER 3: GENERAL DISCUSSION

The goal of the current studies was to investigate the mechanisms and moderators of Situational Attribution Training in reducing automatic stereotype activation. I proposed four hypotheses related to the effects of training on spontaneous situational and trait inferences, the automaticity of the effects of training, and individual differences that may moderate the effectiveness of training.

Hypotheses Review

Hypothesis 1. *Situational Attribution Training increases the tendency to make spontaneous situational inferences for negative stereotypic behaviors performed by African American men.* The Study 1 findings did not support this hypothesis. Participants who were trained to consider situational explanations of negative Black stereotypic behaviors showed automatic activation of SSIs; but so did control participants. These findings are inconsistent with previous research (Stewart et al., 2010), which found evidence of SSI activation only in the training but not in the control condition. However, those data were preliminary and analyses were conducted on a small sample size (N=18), with the overall interaction between probe type and condition not being significant. Thus, the present data suggest caution in interpreting the prior research that showed SSI activation after Situational Attribution Training.

Hypothesis 2. *Participants who complete Situational Attribution Training should show reduced stereotype activation, as suggested by decreased STI activation for negative stereotypic traits.* Study 2 findings failed to replicate entirely the stereotype reduction effects of training found in previous research (Stewart et al., 2010) as well as to show automatic stereotype activation in the control conditions. Regression analyses revealed that only participants high in

NFC showed reliable stereotyping reduction effects after completing Situational Attribution Training. This partial replication may be due to the task used to measure automatic stereotype activation – the probe recognition task – a task rarely used in prejudice reduction research.

Hypothesis 3. *If the effects of Situational Attribution Training have become automatic, participants should show reduced automatic stereotype activation regardless of whether they are cognitively loaded or not when completing the probe recognition task.* As discussed above, reliable stereotype reduction effects were only found for participants who were high in NFC. Contrary to the third hypothesis, these stereotype reduction effects were not maintained under conditions of high cognitive load. This finding suggests that taking into account situations when being exposed to negative Black stereotypic behaviors has not become automatic after Situational Attribution Training.

Hypothesis 4. *Individual differences such as explicit prejudice, motivations to control prejudice, NFC and PNS should moderate the effects of training on automatic stereotype activation.* Consistent with this hypothesis, moderation analyses revealed that Situational Attribution Training worked best for individuals who were initially higher in NFC. Also, training had the opposite effect for individuals high in modern racism who showed an increase in implicit bias. No moderation effects were found for motivations to control prejudice and PNS.

Overall Implications of Findings

There are several explanations and implications of current findings that are relevant to the implicit bias reduction literature. I will discuss three of these issues in detail. First, I discuss the methodological limitations of the current studies and their implications for interpreting the current findings. Second, I discuss a possible theoretical model of the mechanism underlying training success, which may account for the current findings. Third, I discuss findings related to

individual differences moderating training effects, as well as the implication of these findings for designing successful anti-bias interventions. It should be noted that the term “training success” refers to this technique’s effectiveness as documented in previous research (Stewart et al., 2010), as well as in the current research for some participants (high in NFC).

First, the dependent measures used in the second study may not have been appropriate to answer our specific research questions. Using an adapted version of the probe recognition task to measure implicit bias in the form of spontaneous trait inferences, I was not able to replicate the Stewart et al. (2010) findings, which showed that training participants exhibited less stereotype activation compared to controls. A second indication that the probe recognition task in Study 2 did not assess implicit racial bias as measured in the previous literature is that I was not able to replicate the IMS/EMS findings found in racial bias (Devine et al., 2002) and gender bias (Latu et al., in press) research. In fact, using the probe recognition task, I found the opposite pattern of findings from studies that used sequential priming tasks or the IAT. Unfortunately, the failure of Study 2 to replicate basic previous findings casts some doubt on the significant findings of the current research. For example, it is hard to have full confidence and meaningfully interpret the finding that training may work best for individuals high in NFC and low in modern racism given that, overall, I did not replicate the stereotyping reduction effects of training. Future studies should be conducted to elucidate whether the current findings are reliable and replicate with other paradigms or whether the findings are unreliable and only due to methodological peculiarities of the current study. I include a thorough description of future research ideas in the next section of this chapter.

Despite its methodological limitations, findings of the current studies suggest some new directions in understanding the mechanism behind Situational Attribution Training. As

mentioned above, these interpretations should be regarded as tentative given the methodological issues discussed.

Training involves teaching participants to make situational attributions for negative stereotypic behaviors performed by Black men. How does this training technique achieve its success? Is it automatizing participants' tendency to make situational inferences or is it a more controlled process in which participants deliberately change their processing goals from understanding the person to understanding the situation in which the person is? Overall, the findings of two studies suggest that training participants in situational attributions may change individuals' processing goals, such that they are more invested in understanding the situation in which the target is, rather than the stable traits of the person. Consistent with Krull and Erickson's (1995) attributional model, this change in processing goals determines individuals to infer situational causes first, thus circumventing the otherwise automatic tendency of inferring trait or dispositional causes.

There are three main findings from the current studies that suggest that Situational Attribution Training may achieve its success through a conscious process of changing processing goals. First, training did not increase the tendency to make SSIs above and beyond comparable control training. Although the current data do not address the effects of training on intentional inferences, future studies may show that training has a significant effect on such intentional situational inferences. Second, the automatic stereotype reduction effects of training for high NFC participants were only observed when participants were not cognitively loaded while completing the dependent task. If training effects disappear when participants are engaged in a competing cognitive task, the tendency to take into account situational factors (thus correcting for inflated trait attributions) has not yet become automatic. Thus, it may be that a consciously

controlled process is responsible for the Situational Attribution Training effects. Last, explicit prejudice moderated the effects of training on automatic stereotype activation, such that training worked best for participants who were low in modern racism. The finding that a conscious, explicit attitude makes participants more likely to respond positively to training suggests that its effects may be consciously controlled.

Present findings suggest that a conscious process of changing processing goals may account for the effectiveness of Situational Attribution Training in decreasing stereotype activation. But how would this process look like? How does a conscious process (such as changing goals) permeate to the automatic level such that automatic stereotype activation is decreased? Figure 3 presents a summary of this proposed process. Consistent with Krull and Erickson's three-stage attribution model (1995), Situational Attribution Training may be changing individuals' processing focus from understanding the person to understanding the situation. In turn, this change in focus may sabotage trait inferences, thus reducing negative trait – racial group associations. This hypothesis should be investigated in future research.

One interesting option would be to measure both trait associations and situational inferences following training. This strategy would offer the opportunity to conduct a mediation analysis to investigate if increases in situational inferences mediate the relationship between training and reduced trait activation. This goal may be difficult to achieve, however, given that the concomitant activation of both situational and trait inferences may be hard to measure. In fact, previous research has never documented the co-occurrence of trait and situational inferences for the same person (in a within subjects design), only for the same behavior (in a between subjects design; Ham & Vonk, 2006). Attempting to document the activation of both trait and situational inferences for the same perceiver has its risks, because their activation may

prove to be mutually exclusive. In addition, once an inference (situational or trait) is activated during its assessment, it can interfere with and bias the assessment of other inferences.

Other studies also looked at how conscious goals reduce automatic stereotype activation. For example, Stewart and Payne (2008) had participants complete a weapons identification task in which they had to decide whether objects presented on the screen were tools or weapons, after being primed with photos of Black and White men. In previous research using this task (Payne, 2001), participants showed evidence of automatic stereotype activation, as they were quicker to decide that an object was a gun after being primed with a Black compared to a White photo, and they were more likely to incorrectly classify a tool as a gun after being primed with a Black compared to a White photo. In Stewart and Payne, while completing the weapons identification task, participants were told to think non-stereotypic thoughts (thinking “safe” when seeing a Black man). Compared to controls, participants who engaged in this conscious control strategy showed reduced automatic stereotype activation. Using a similar conscious strategy, Kawakami and her colleagues (2002) found reduced stereotype activation after participants repeatedly said NO to traits stereotypic of Blacks, skinheads, and elderly people.

However, one risk of such techniques is that conscious thought suppression may ironically increase the frequency of the suppressed thought, a phenomenon called the rebound effect. Previous research revealed that stereotype suppression can actually lead to an increase in stereotyping, due to a rebound effect (Macrae et al., 1994, Payne, Lambert, & Jacoby, 2002). The Situational Attribution Training technique avoids such rebound effects, as it does not ask participants to suppress stereotypes. In fact, the training task is not even presented to participants as being related to racial stereotyping. Instead, participants are trained to think of situational

explanations, and this novel attributional strategy led to reduced stereotype activation. Thus, the ironic effects of suppression are likely to be reduced.

Finally, another major contribution that the current research makes to the implicit bias reduction literature is unraveling the role of individual differences in response to training. With few exceptions, there are no studies that systematically investigated how individual differences relate to stereotyping reduction interventions. The current research revealed that individuals high in NFC and low in explicit prejudice showed the most pronounced benefits from Situational Attribution Training. How can this training technique be adapted to address individuals who are low in NFC and high in modern racism? Individuals low in NFC do not enjoy and seek to process novel information. Thus, these individuals may benefit from a training technique which automatizes certain processing strategies – a tactic which would diminish their processing needs. For example, it may be that repeated, shorter training sessions may increase automatic situational inferences and, as a consequence, reduce stereotyping. Individuals high in modern racism explicitly believe that racial prejudice and discrimination are no longer a problem. As such, it is possible that they showed more reactance to training, guessing that it was designed to reduce racial stereotyping. In the future, it may be that reducing demand characteristics may enhance training effects for participants who are high in modern racism, by reducing reactance. One idea would be to train participants in making situational attribution not only for Black but also for White stereotypic behaviors. This strategy would also be of theoretical interest, because it would allow us to investigate whether teaching participants a general attributional strategy, not necessarily specifically related to Black stereotypes, would reduce automatic stereotype activation.

Future Directions

In the current studies I answered some questions about the mechanisms and moderators of Situational Attribution Training; current findings, however, also raised as many interesting questions, which inform plans for future research. In future studies I plan to make several modifications to the current paradigm in order to test the mechanism and moderators of Situational Attribution Training.

First, future research should investigate the effects of Situational Attribution Training on intentional trait and situational inferences using one of the existing research paradigms. For example, one of the commonly used inference research paradigms dates back to Jones and Davis' (1965) work on the correspondence bias and involves the presentation of several behaviors. In that paradigm, the participants' task is to indicate the extent to which certain traits or situational properties account for those behaviors. In a more recent version of this paradigm, Gilbert et al. (1988) had participants watch a videotape of a job interview with a woman who was behaving in an anxious manner. Participants were asked to rate on a 9-point Likert scale how anxious the woman was (trait factor) as well as how anxiety provoking they thought the interview was (situational factor). A similar methodology could be employed to investigate the effects of Situational Attribution Training on intentional trait and situational inferences. Following training, participants would be exposed to several negative Black-stereotypic behaviors not presented during training, in written or video format. Afterwards, participants' intentional inferences of those behaviors would be assessed. It may be, however, that participants who completed training would deliberately inflate their intentional situational attributions because of demand characteristics associated with training. To minimize these effects, the research paradigm could be adapted to require an open-ended response from participants. Their

task would be to write a paragraph explaining the causes of the person's behavior. The output would be analyzed using content analysis, by counting the number of trait and situational explanations employed by participants. This strategy may not fully eliminate demand characteristics, but it would diminish their impact.

Overall, further investigating the effects of Situational Attribution Training on both intentional and automatic inferences (using alternative research methods and designs) would bring insight into the processes underlying the previously documented effects of Situational Attribution Training. Is Situational Attribution Training changing participants' inference goals, such that they are more invested in understanding situations rather than traits? If this is the case, we should see that training increases intentional situational inferences, but not spontaneous situational inferences. The other possibility may be that participants' goal remains that of understanding the person and not the situation, and training is automatizing the tendency to take into account the situation, in addition to inferring traits automatically. If this is the case, we should see that training increases spontaneous, but not intentional situational inferences.

Second, I would make several modifications to the second study of the current research. To start with, I would not use the probe recognition task to measure automatic stereotyping; instead, I would replicate previous findings from Stewart and colleagues (2010) using a sequential priming task such as the lexical decision task (Wittenbrink et al., 1997) or the IAT (Greenwald et al., 1998). These tasks have been commonly used in stereotype reduction research (see Blair, 2001 for a review) and are thus sensitive enough to capture variations of automatic associations. Moreover, tasks such as the lexical decision task may be even more appropriate and sensitive than the person categorization task in measuring stereotype activation. In the person categorization task, participants categorize Black and White faces after being primed with Black

stereotypic words. The underlying idea is that stereotypic traits would activate the racial stereotype, which would facilitate the categorization of Black versus White faces. In the lexical decision task, the order of the trait and photo is reversed, such that participants classify traits as positive or negative after being primed with Black and White faces. I would expect this measure to be more sensitive than the person categorization task because the exposure to a group exemplar (i.e., the Black photo) would be more likely to activate certain stereotypic traits (e.g., lazy) compared to the reversed order. This presentation order would also map more closely onto the theoretical understanding of stereotyping in which the group activates the stereotypic trait and not vice-versa.

Also, using the lexical decision task as a measure of stereotype activation, I would attempt to replicate the moderating effects found in the current study. I would measure individual differences such as explicit prejudice, motivations to control prejudice, NFC, and PNS before participants complete Situational Attribution Training. Hopefully, findings would not only replicate those of the current study, but they would also strengthen the marginal effects found for NFC in the current study. Additionally, I would also investigate the moderating role of attributional complexity in moderating responses to training. Attributional complexity (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986) refers to the extent to which individuals use complex schemas in determining behavioral causes. This individual difference is strongly related to individuals' tendency to use situational explanations in determining the cause of others' behaviors, as suggested by Fletcher and colleagues' multi-construct definition of attributional complexity. Three of these constructs measure situational explanations: awareness of how interaction with other people influence behaviors (e.g., "I think very little about the different ways that people influence each other."), awareness of abstract, distal causes that may determine

behaviors (e.g., “I think a lot about the influence that society has on other people.”), as well as awareness of external causes from the past which may determine behaviors (e.g., When I analyze a person’s behavior I often find the causes from a chain that goes back in time, sometimes for years.”). Attributional complexity correlates with several variables relevant to the proposed studies. For example, it is significantly related to NFC (Fletcher et al.), perspective taking and empathy (Joireman, 2004) and explicit stereotyping of minimal groups (Schaller, Boyd, Yohannes, O’Brien, 1995). Based on these findings, I hypothesize that attributional complexity will moderate Situational Attribution Training effects, such that individuals high in attributional complexity will have an easier time completing the training and will also gain the most benefits from training in terms of implicit bias reduction.

Finally, I would use a cognitive load task while participants are completing the intentional situational inference task, to investigate whether training is automatizing the tendency to make intentional situational inferences. In the current study, the cognitive load task accompanied the probe recognition task, which assesses the relatively automatic process of trait inferences. Thus, it was difficult to assess whether a process that was already automatic has become even more automatic after training. By adding cognitive load to an explicit, intentional situational inference task, I would be able to assess more precisely whether the process of inferring situational causes has increased in automaticity after Situational Attribution Training.

Training Applications

Once the effectiveness of Situational Attribution Training is documented in the controlled, laboratory environment, it would be interesting to design several real-world applications of this technique. For example, groups other than college students may be especially

likely to benefit from Situational Attribution Training. Training could reduce teachers' racial and gender stereotypes and thus reduce the likelihood of stereotype threat for minority students. Similarly, training police officers in situational attributions may decrease racially biased shooting errors previously documented with the shoot/don't shoot paradigm (Correll et al., 2002).

Before Situational Attribution Training is implemented in applied settings, several steps need to be taken inside the lab to ensure the maximum efficiency of this technique. For example, future studies should establish if shorter (possibly repeated) training sessions may be more effective than the current version of training. Also, future studies should investigate whether teaching participants to consider situational factors of African American men's behaviors would also decrease implicit bias towards other groups on which training was not conducted. This finding would be consistent with the idea that Situational Attribution Training is teaching participants to automatize a novel critical thinking strategy that determines them to seek situational explanations for others' behaviors. In addition, future research should be conducted to establish the durability of its effectiveness. Theoretically, the current training technique has less of a chance of a rebound effect because of its indirect nature and lack of stereotype negation. However, future empirical work needs to be conducted to test this assumption and to investigate whether training effects are maintained in 24 hours, a week, and even longer.

Conclusion

In the stereotyping reduction literature it is important to document how and for whom certain interventions are effective. The current studies bring us a step closer to answering these questions related to Situational Attribution Training. With some methodological caveats, the mechanism underlying the success of training seems to be a consciously controlled one – that of

taking into account situational factors, suggesting that this training technique circumvents the well-documented tendency to engage in the ultimate attribution error. Of theoretical and practical importance, Situational Attribution Training is more efficient for individuals low in modern racism and high in NFC. These findings can inform future research, by designing stereotyping reduction interventions that are customized for individuals based on their individual differences.

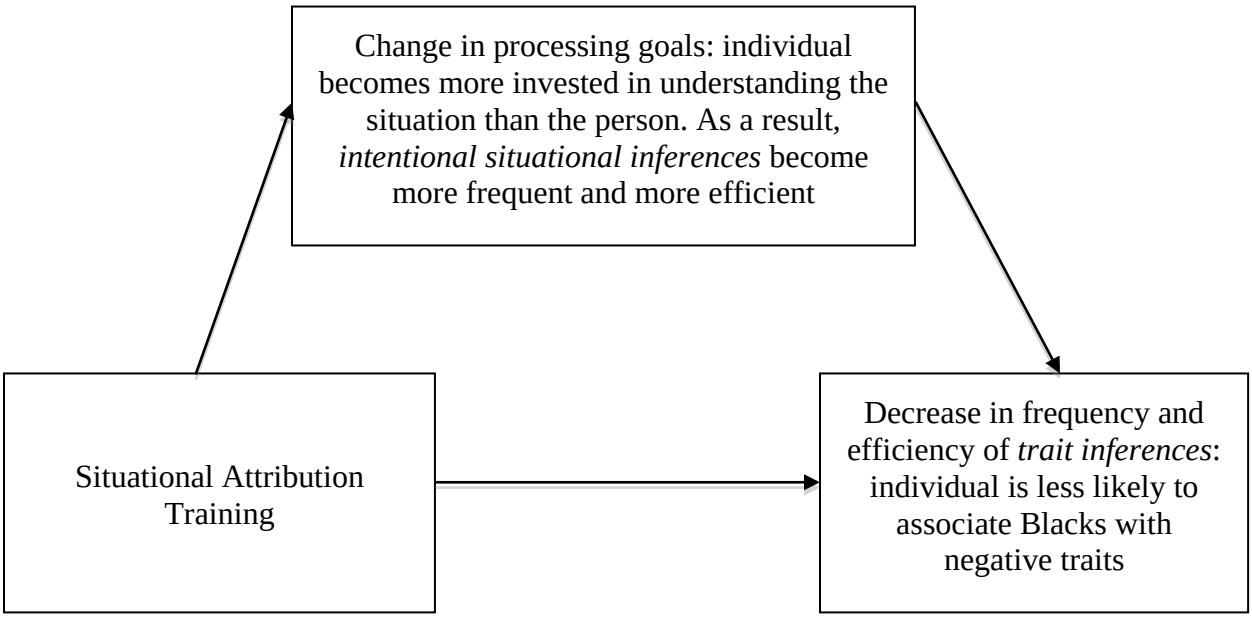


Figure 3. The Mechanism Behind Situational Attribution Training

ENDNOTES

[1] I also pretested the load manipulation in conjunction with the person categorization task, which was used in previous research (Stewart et al., 2010) to measure automatic stereotype activation following Situational Attribution Training. In this paradigm, participants' task was to categorize Black and White photographs by race as quickly as possible, after being exposed to a trait for 240ms. Of interest were eight negative traits that were stereotypic of Black men, and that were not presented during the training phase. To the extent that participants spontaneously activate negative stereotypes of Blacks, they should be faster to categorize a Black photo compared to a White photo, after being primed with a negative stereotypic trait. In the load pretest, 33 participants completed the person categorization task while being randomly assigned to either a low or high cognitive load condition. The load manipulation was adapted from several load tasks used in previous research (e.g., Bodner & Stalinski, 2008, Stewart et al., 2003, Wigboldus et al., 2004; see Table 1 for a summary of cognitive load manipulations in the literature). In the low load condition, 16 participants were presented with 16 one-digit numbers for 1,000ms, at random intervals between 5 to 9 trials of the person categorization task. Participants' task was to remember and recall these numbers in writing, before a new number was presented. In the high load condition, 17 participants memorized and recalled a total of 16 six and seven-digit numbers that were presented for 6,000 and 7,000 ms, respectively. Presumably, rehearsing five and six-digit numbers while completing the categorization task would lead to significantly higher cognitive load compared to rehearsing one-digit numbers. The results confirmed this hypothesis: participants were overall slower to categorize photos in the high load condition (M = 627.03) compared to the low load condition (M = 550.13), $t(31) = 2.02$, $p = .05$, Cohen's $d = .72$. However, of interest to our research goals, I was not able to see evidence of automatic stereotype

activation in either condition, as participants were not faster to categorize Black compared to White photos after being primed with a Black negative trait either in the low load condition, $F(15) = .71$, $p = .41$, $\eta^2 = .04$, or the high load condition, $F(16) = .79$, $p = .39$, $\eta^2 = .05$. As such, the probe recognition task was used instead, because previous research (Wigboldus et al. 2003, Stewart et al., 2003) has successfully used cognitive load manipulations in conjunction with this task to measure the automaticity of stereotypic trait inferences.

[2] This effect was not significant in equivalent analyses using the difference score between control and trait probes for White photo trials.

REFERENCES

Allen, T.J. Sherman, J.W., & Klauer, K.C. (2010). Social context and the self-regulation of implicit bias. *Group Processes and Intergroup Relations, 13,* 137-149.

Allport, G. W. (1954/1979). *The nature of prejudice.* Perseus Books.

Amodio, D.M. (2008). The social neuroscience of intergroup relations. *European Review of Social Psychology, 19,* 1-54.

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,* 652-661.

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 responses and self-report. *Journal of Personality and Social Psychology, 84,* 738-753.

Amodio, D. M., & Mendoza, S. A. (in press). Implicit intergroup bias. To appear in B. Gawronski and B. K. Payne (Eds.) *Handbook of Implicit Social Cognition.* New York: Guilford.

Bargh, J. A. (1994). The four horseman of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of Social Cognition* (2nd ed.). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Pietromonaeo, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology, 43,* 437-449.

Batson, C.D., Polycarpou, M.P., Harmon-Jones, E., Imhoff, H.J., Mitchener, E.C., Bednar, et al. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72,* 105-118.

Bessenoff, G.R. & Sherman, J.W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition,18,* 329-353.

Blair, I.V. & Banaji, M.R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70,* 1142-1163.

Blair, I.V., Ma, J.E., & Lenton, A.P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81,* 828-841.

Bodenhausen, G.V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 1,* 319-322.

Bodner, G.E. & Salinski, S.M. (2008). Masked repetition priming and proportion effects under cognitive load. *Canadian Journal of Experimental Psychology, 62,* 127-131.

Bogardus, E. S. (1925). Measuring social distances. *Journal of Applied Sociology, 9,* 299-308.

Brescoll, V.L. & Uhlmann, E.L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science, 19,* 268-275.

Brigham, J.C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23*, 1933-1967.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131

Cacioppo, J.T., Petty, R.E., & Kao, C.F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Cialdini, R.B., Trost, M.R., & Newsom, J.T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology, 69*, 318-328.

Coats, S., Latu, I.M., & Haydel, L.A. (2007). The facilitative effects of evaluative fit on social categorization. *Current Research in Social Psychology, 12*, 54-67.

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*, 1314-1329.

Crawford, M. T., & Skowronski, J. J. (1998). When motivated thought leads to heightened bias: High need for cognition can enhance the impact of stereotypes on memory. *Personality and Social Psychology Bulletin, 24*, 1075-1089.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 801-814.

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and fragile: A new look at the automaticity of negation processing. *Experimental Psychology, 56,* 434–446.

Devine, P. G. (1989). Prejudice and stereotypes: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5-18.

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,* 835-848.

Dijksterhuis, A., Aarts, H., Bargh, J.A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology, 36,* 531-544.

Duff, K. J., & Newman, L. S. (1997). Individual differences in the spontaneous construal of behavior: Ideocentrism and the automatization of the trait inference process. *Social Cognition, 15,* 217– 241.

Duncan, B.L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology, 37,* 621-634.

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

Dovidio, J. F., ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., et al. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin, 30*, 1537–1549.

Dovidio JF, Kawakami K, Gaertner SL. (2002). Implicit and explicit prejudice and interracial interactions. *Journal of Personality and Social Psychology, 82*, 62–68.

Fazio, R.H. & Olson, M.A (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Fazio, R.H.. & Towels-Schwen, T. (1999). The MODE model of attitude-behavior processes. In *Dual Process Theories in Social Psychology.* Eds. S Chaiken & Y. Troppe, pp 97-116, New York: Guilford.

Fiske, S.T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48*, 621–628.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford Press.

Fletcher, G.J.O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. *Journal of Personality and Social Psychology, 51,* 875-884.

Galinsky, A.D. & Moskowitz, G.B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78,* 708-724.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692-731.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology, 44,* 370-377.

Gawronski, B., LeBel, E.P, & Peters, K.R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2,* 181-193.

Gilbert, G.M. (1951). Stereotype persistence and change among college students. *Journal of Abnormal and Social Psychology, 46,* 245-254.

Gilbert, D. T., & Gill, M. J. (2000). The momentary realist. *Psychological Science, 11,* 394-398.

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational Behavior and Human Decision Processes, 88,* 430-444.

Gilbert, D.T. & Hixon, J.G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509-517.

Gilbert, D.T. & Malone, P.S. (1995). The correspondence bias, *Psychological Bulletin, 117,* 21-38.

Gilbert, D. T., Pelham, B. W. & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733-740.

Greenberg, J. & Rosenfield, D. (1979). Whites' ethnocentrism and their attributions for the behavior of Blacks: A motivational bias. *Journal of Personality, 47,* 643-657.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition. *Psychological Review, 102,* 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition. The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97,* 17–41

Ham, J & Vonk, R. (2003). Smart and easy: Co-occuring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology, 39,* 434-447.

Heitland, K., & Bohner, G. (2009). Reducing prejudice via cognitive dissonance: Individual differences in preference for consistency moderate the effects of counterattitudinal advocacy. *Social Influence*.

Henry, P.J. & Hardin, C.D. (2006). The contact hypothesis: Status bias in the reduction of implicit prejudice in the United States and Lebanon. *Psychological Science, 17,* 862-868.

Hewstone, M. (1990). The ‘ultimate attribution error’? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology, 20,* 311-335.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology, 36,* 660-676.

Joireman, J. (2004). Relationships between attributional complexity and empathy. *Individual Differences Research, 2,* 197-202.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in social perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology:* Vol. 2 (pp. 219–266). New York: Academic Press.

Judd, C., Kenny, D., and McClelland, H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6,* 115-134.

Juni, S, Brannon, R. & Roth, M.M. (1988). Sexual and racial discrimination in service-seeking interactions: A field study in fast food and commercial establishments. *Psychological Reports, 63,* 71-76.

Karlins, M., Coffman, T. L., &Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology, 13,* 1-16.

Katz, D & Braly, K.W. (1933).Racial stereotypes of one-hundred college students. *Journal of Abnormal and Social Psychology, 28,* 280-290.

Kawakami, K., Dion, K.L., & Dovidio, J.F. (1998) Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin*, 24, 407-416.

Kawakami, K., Dovidio, J.F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation, *Journal of Personality and Social Psychology, 78,* 871-888.

Keltner, D., Gruenfeld, D.H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110,* 265–284.

Kempf, K.L. & Austin, R.L. (1986). Older and more recent evidence on racial discrimination in sentencing. *Journal of Quantitative Criminology. 2,* 29-48.

Kinder, D. R., & Sanders, L. M. (1996). *Divided by color*. Chicago: University of Chicago Press.

Koole, S.L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80,* 669–685.

Krull, D.S. & Erikson, D.J. (1995). Inferential hopscotch: How people draw social inferences from behavior. *Current Directions in Psychological Science, 4,* 35-38.

Lalwani, A.K. (2009). The Distinct Influence of Cognitive Busyness and Need for Closure on Cultural Differences in Socially Desirable Responding. *Journal of Consumer Research 36*, 305-316.

Latu, I.M., Stewart, T.L., Myers, A.C., Lisco, C.G., Estes, S.B., & Donohue, D. (in press). What we "say" and what we "think" about female managers: Explicit versus implicit associations of women with success. *Psychology of Women Quarterly.*

Levy, S.R. (1999). Reducing prejudice: Lessons from social-cognitive factors underlying perceiver differences in prejudice. *Journal of Social Issues, 55,* 745-765.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81,* 842– 855.

Lupfer, M.B., Clark, L.F., Church, M., DePaola, S.J., & McDonald, C.D. Do people make situational as well as trait inferences spontaneously? Unpublished manuscript.

Macrae, C.N., Bodenhausen, G.V., Milne, A.B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67,* 808-817.

Madon, S, Guyll, M. Aboufadel, K, Montiel, E, Smith, A, Palumbo, P. et al. (2001). Ethnic and national stereotypes: The Princeton Trilogy revisited and revised. *Personality and Social Psychology Bulletin, 27,* 996-1010.

McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin, 9,* 551–558.

McConnel, A.R. & Leibold, J.M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435-442.

McKoon, G. & Ratcliff, R. (1989). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326-338.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory, structure, and mental processes. *Science*, *192*, 27-33.

Mikulincer, M. Birnbaum, G., Woddis, D., & Nachmias, O. (2000). Stress and accessibility of proximity-related thoughts: Exploring the normative and intraindividual components of attachment theory. *Journal of Personality and Social Psychology, 78*, 509-523.

Monteith, M.J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 72,* 420-434.

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214-255.

Neuberg, S.L. & Newsom, J.T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology, 65, 113-131.*

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*, 14-29.

Osborne, R. E., & Gilbert, D. T. (1990). *The preoccupational hazards of social life.* Unpublished manuscript, University of Texas at Austin.

Osterhouse, R. A., & Brock, T. C. (1970). Distraction increases yielding to propaganda by inhibiting counterarguing. *Journal of Personality and Social Psychology*, *15*, 344-358.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181-192.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277-293.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38,* 384-396

Paulhus, D.L., (1984). Two component models of social desirable responding. *Journal of Personality and Social Psychology, 46,* 598– 609.

Pendry, L. F., & Macrae, C. N. (1999). Cognitive load and person memory: The role of perceived group variability. *European Journal of Social Psychology, 29*, 925–942.

Pettigrew, T.F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5,* 461-476.

Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90,* 751-783.

Petty, R.E., Wells, G.L., Brock, T.C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34,* 874-884.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby J. C., Gore, J. C., Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729-738.

Plant, E.A. & Devine, P.G. (1998). Internal and external motivations to control prejudice. *Journal of Personality and Social Psychology, 75,* 811-832.

Rankin, R.E. & Campbell, D.T (1955). Galvanic skin responses to Negro and white experimenters. *Journal of Abnormal Psychology, 51,* 30-33.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In Berkowitz (Ed.), *Advances in experimental social psychology* (Vol 10, pp. 173-220). San Diego, CA: Academic Press.

Rudman, L.A. & Ashmore, R.D. (2007). Discrimination and the Implicit Association Test, *Group Processes and Intergroup Relations, 3,* 359-372.

Rudman, L.A. & Kilianski, S.E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26,* 1315-1328.

Schaller, M., Asp, C. H., Rosell, M. C., & Heim, S. J. (1996). Training in statistical reasoning inhibits the formation of erroneous group stereotypes. *Personality and Social Psychology Bulletin, 22,* 829-844.

Schaller, M., Boyd, C., Yohannes, J., & O'Brien, M. (1995). The prejudiced personality revisited: Personal need for structure and formation of erroneous group stereotypes. *Journal of Personality and Social Psychology, 68,* 544-555.

Sherman, J.W. & Frost, L.A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin, 26,* 26-34.

Sherman, J.W., Lee, A.Y, Bessenoff, G.R., & Frost, L.A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75,* 589-606.

Sinclair, L.,& Kunda, Z.(1999).Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885–904.

Sinclair, S., Lowery, B.S., Hardin, C.D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative orientation. *Journal of Personality and Social Psychology, 89*, 583-592.

Skitka, L.J., Mullen, E. Griffin, T., Hutchinson, S. & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction? Understanding ideological differences in explanations for social problems, *Journal of Personality and Social Psychology, 87*, 470-487.

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

Stewart, T.L., Latu, I.M., Kawakami, K., & Myers, A.C. (2010). Consider the situation: Reducing automatic stereotyping through situational attribution training. *Journal of Experimental Social Psychology, 46*, 221-225.

Stewart, B.D., & Payne, B.K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin, 34*, 1332-1345.

Stewart, T.L., Weeks, M., & Lupfer, M.B. (2003), Spontaneous stereotyping: A matter of prejudice? *Social Cognition, 21*, 263-298.

Tajfel, H. & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole.

Taylor, S. E., Fiske, S. T., Etcoff, N. J., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36, 778-793.*

Uleman, J.S., Hon, A., Roman, R.J., & Moskowitz, G.B. (1996a). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin, 22,* 377-394.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996b). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology:* Vol. 28 (pp. 211–279). San Diego, CA: Academic Press.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73*, 941-959.

Vescio, T.K., & Sechrist, G.B., & Paolucci, M.P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology, 33,* 455-472.

Van den Bos, K., Peters, S. L., Bobocel, D. R., & Ybema, J. F. (2006). On preferences and doing the right thing: Satisfaction with advantageous inequity when cognitive processing is limited. *Journal of Experimental Social Psychology, 42,* 273-289.

Wigboldus, D.H.J., Sherman, J.W., Franzese, H.L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition, 22,* 292-309.

Wittenbrink, B., Gist, P.L., & Hilton, J.L. (1997). Structural properties of stereotypic knowledge and their influences on the construal of social situations. *Journal of Personality and Social Psychology, 72,* 526-543.

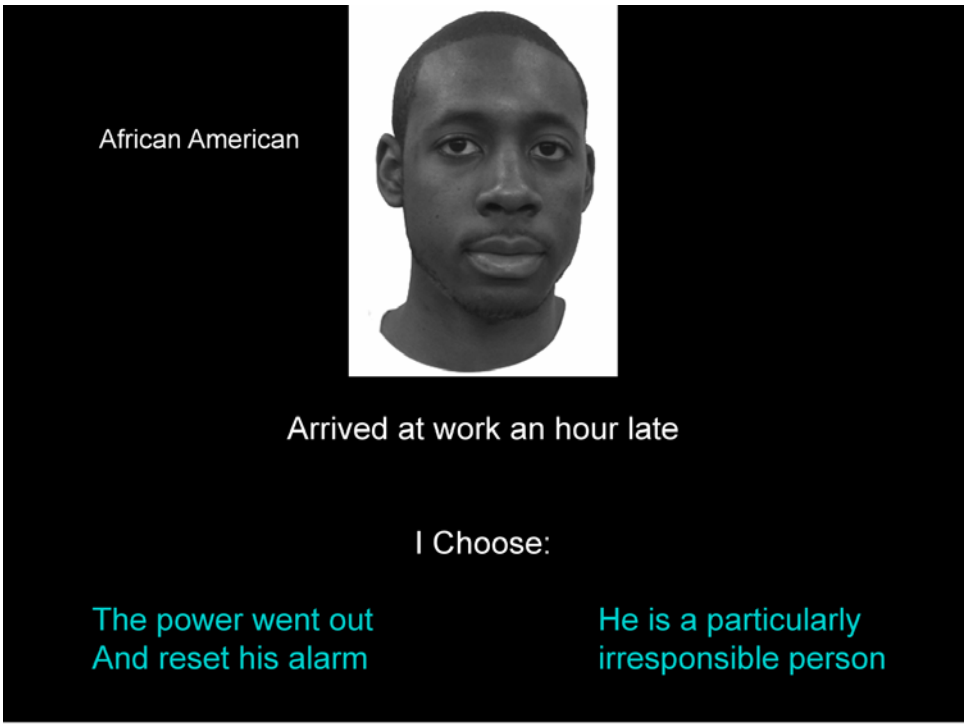
Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Wittenbrink, B., Judd, C.M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815-827.

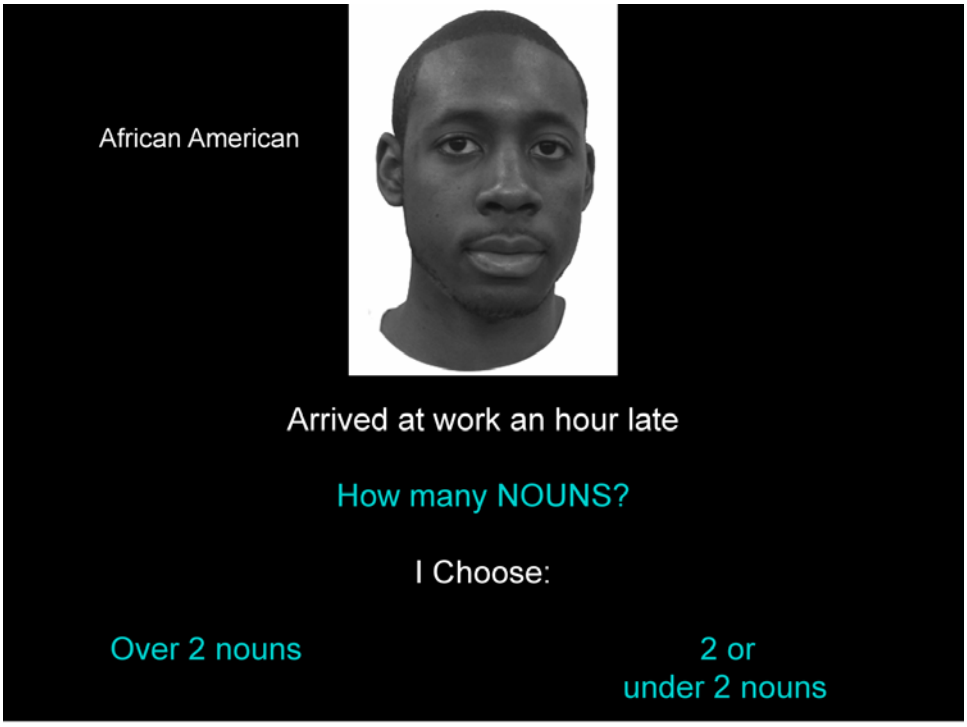
Ziegert, J.C. & Hanges, P.J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology, 90,* 553-362

APPENDIX A - CONDITIONS

Example of a Typical Screen in the Situational Attribution Training Condition



Example of a Typical Screen in the Grammar Control Condition



APPENDIX B - SCALES

Social Distance Scale

I would be willing to have a Black American person as my:

	STRONGLY DISAGREE								STRONGLY AGREE
Good Friend	1	2	3	4	5	6	7	8	9
Next Door Neighbor	1	2	3	4	5	6	7	8	9
Co-worker	1	2	3	4	5	6	7	8	9
Roommate	1	2	3	4	5	6	7	8	9
Child's Friend	1	2	3	4	5	6	7	8	9
Sibling's spouse	1	2	3	4	5	6	7	8	9
Romantic Date	1	2	3	4	5	6	7	8	9
Family physician	1	2	3	4	5	6	7	8	9
U.S. President	1	2	3	4	5	6	7	8	9
Governor	1	2	3	4	5	6	7	8	9
Wife or Husband	1	2	3	4	5	6	7	8	9
Child's teacher	1	2	3	4	5	6	7	8	9
Dance partner	1	2	3	4	5	6	7	8	9
Fellow church or Social club member	1	2	3	4	5	6	7	8	9

Attitude towards Blacks/Modern Racism Scale

If I had a chance to introduce Black visitors to my friends and neighbors, I would be pleased to do so.

Some Blacks are so touchy about race that it is difficult to get along with them.

The federal government should take decisive steps to override the injustices Blacks suffer at the hand of local authorities.

I would rather not have Blacks live in the same apartment building I live in.

I get very upset when I hear a White make a prejudicial remark about Blacks.

Blacks and Whites are inherently equal.

It is likely that Blacks will bring violence into neighborhoods when they move in.

Discrimination against Blacks is no longer a problem in the United States.

Blacks are too demanding in their push for equal rights.

Blacks should not push themselves where they are not wanted.

I would not mind at all if a Black family with about the same income and education moved in next door.

I enjoy a funny racial joke, even if some people might find it offensive.

Racial integration (of schools, businesses, residences, etc.) has benefited both Whites and Blacks.

I would probably feel somewhat self-conscious dancing with a Black person in a public place.

I favor open housing laws that allow more racial integration of neighborhoods.

Whites should support Blacks in their struggle against discrimination and segregation.

Generally, Blacks are not as smart as Whites.

Over the past few years, the government and news media have shown more respect for Blacks than they deserve.

Over the past few years, Blacks have gotten more economically than they deserve.

Blacks have more influence on school desegregation plans than they deserve.

If a Black were put in charge of me, I would not mind taking advice and direction from him or her.

I think that Black people look more similar to each other than White people do.

Interracial marriage should be discouraged to avoid the “who-am-I?” confusion, which the children feel.

It would not bother me if my new roommate was Black.

I worry that in the next few years I may be denied my application for a job or promotion because of preferential treatment given to minority group members.

Internal/External Motivation to Control Prejudice Scale

Because of today's PC (politically correct) standards I try to appear nonprejudiced toward Black people.

I try to hide any negative thoughts about Black people in order to avoid negative reactions from others.

If I acted prejudiced toward Black people, I would be concerned that others would be angry with me.

I attempt to appear nonprejudiced toward Black people in order to avoid disapproval from others.

I try to act nonprejudiced toward Black people because of pressure from others.

I attempt to act in nonprejudiced ways toward Black people because it is personally important to me.

According to my personal values, using stereotypes about Black people is OK.

I am personally motivated by my beliefs to be nonprejudiced toward Black people.

Because of my personal values, I believe that using stereotypes about Black people is wrong.

Being nonprejudiced toward Black people is important to my self-concept.

The need for cognition scale

I would prefer complex to simple problems.

I like to have the responsibility of handling a situation that requires a lot of thinking.

Thinking is not my idea of fun.

I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

I try to anticipate and avoid situations where there is a likely chance that I will have to think in depth about something.

I find satisfaction in deliberating hard and for long hours.

I only think as hard as I have to.

I prefer to think about small, daily projects as opposed to long-term ones.

I like tasks that require little thought once I've learned then.

The idea of relying on thought to make my way to the top appeals to me.

I really enjoy a task that involves coming up with new solutions to problems.

Learning new ways to think doesn't excite me very much.

I prefer my life to be filled with puzzles that I must solve.

The notion of thinking abstractly is appealing to me.

I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

I feel relief rather than satisfaction after completing a task that requires a lot of mental effort.

It’s enough for me that something gets the job done; I don’t care how and why it works.

I usually end up deliberating about issues even when they do not affect me personally.

Personal Need for Structure

It upsets me to go into a situation without knowing what I can expect from it.

I'm not bothered by things that interrupt my daily routine.

I enjoy having a clear and structured mode of life.

I like to have a place for everything and everything in its place.

I enjoy being spontaneous.

I don't like situations that are uncertain.

I hate to change my plans at the last minute.

I hate to be with people that are unpredictable.

I find that a consistent routine enables me to enjoy life more.

I enjoy the exhilaration of being in unpredictable situations.

I become uncomfortable when the rules in a situation are not clear.