From Learning to Doing: The Effects of Educating Individuals on the Pervasiveness of
Bias

Jennifer Alana Joy-Gaba
Cumberland, MD

B. S., Psychology, May 2005
B. S., Mathematics, May 2005
M. A., University of Virginia, January 2008

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
August, 2011

Brian A. Nosek, Ph.D. (Chair)

Bethany A. Teachman, Ph.D.

Timothy D. Wilson, Ph.D.

M. Norman Oliver, Ph.D., M.D.

Abstract

Despite many trying to be egalitarian in social judgment, discrimination still occurs. One reason for the discrepancy between values and behavior may be the “bias blind spot” (Pronin & Kugler, 2007), which suggests that individuals more easily recognize bias in others than they recognize in themselves. Many people believe that they are objective and impartial and, as a consequence, falsely conclude that they are immune to biased judgments based on social group memberships. The purpose of this dissertation was to investigate whether providing experiential education about one’s own biases can shift beliefs about bias and social behavior. Study 1 and 2 examined the immediate and long-term impact on *beliefs about bias* of an interactive lecture about how biases are an ordinary part of human thinking. Study 2 also assessed whether bias education predicts change in social judgments in which biases could be expressed and whether the components of *beliefs about bias* relate to other social attitudes and beliefs. Study 3 and 4 experimentally compared a shortened version of the automatic bias education to already established interventions. Finally, Studies 5-6 explored whether the shortened version of the education was effective. Results revealed that the original automatic bias education, and not the shortened version, was effective in changing people’s *beliefs about bias* in the short and long-term. However, this automatic bias education was not able to influence participants’ automatic and self-reported racial attitudes or judgments in a stereotyping task.

Table of Contents

Abstract.....	ii
Table of Contents.....	iii
Dedication and Acknowledgements.....	vi
Chapter 1: Overview.....	1
1. Egalitarian Values Do Not Guarantee Unbiased Thoughts or Behavior.....	1
1.1 Controlled Processes.....	2
1.2 Automatic Processes.....	2
2. Education and Bias.....	4
2.1 The Bias Blindspot.....	6
2.2 Experiential Education about Unintentional Bias may Change Beliefs about Bias.....	7
2.2.1 Acknowledgement that Automatic Bias Exists.....	7
2.2.2 Recognition that Automatic Bias can Influence Behavior.....	8
2.2.3 Self-Perception of Objectivity.....	8
2.2.4 Motivation to Address Automatic Bias.....	9
3. Automatic Bias Education and Changing Beliefs about Bias May Change Social Judgments and Behavior.....	10
4. Why Education Alone Might Not Change Individuals' Social Judgment and Behavior.....	11
Chapter 2: Overview of Studies.....	12

1. Study 1: Automatic Bias Education Influences People's <i>Beliefs about Bias</i>	13
1.1 Method.....	14
1.2 Results and Discussion.....	19
2. Study 2: Automatic Bias Education is Durable.....	41
2.1 Method.....	48
2.2 Results and Discussion.....	57
3. Study 3: Automatic Bias Education Compared to Other Interventions...	77
3.1 Method.....	80
3.2 Results and Discussion.....	84
4. Studies 4-6: Focused Studies Examining the Shortened Bias Education.....	91
4.1 Method.....	94
4.2 Results and Discussion.....	97
Chapter 3: General Discussion and Conclusions.....	105
1. Summary of Studies.....	105
2. Why the Shortened Bias Education was not as Impactful as the Original Education.....	106
3. Why Some Factors May be More Durable than Others.....	109
4. Automatic Bias Education Differentially Impacts Social Groups.....	111
5. Why Automatic Bias Education was Ineffective at Changing Participants' Attitudes and Judgments.....	112

6. Conclusion.....	113
References.....	116
Appendix.....	125

Dedication and Acknowledgements

I would like to dedicate this dissertation first and foremost to my advisor, Brian Nosek. I am at a loss for words on how to thank you for all you have done for me. Suffice it to say I am deeply grateful to have had such an amazing and wonderful mentor. You had faith in me even during those times when I didn't. It is through that faith that I mustered the confidence and the drive to pull through what sometimes seemed (to me) the impossible. Because of you, I have learned to greatly value transparency. I have also learned the importance of not only doing great science, but in challenging the process to make it better for future generations. I thank you for your patience, guidance, and support that you have given me throughout the years.

I am sincerely grateful to Bethany Teachman, Tim Wilson, and Norm Oliver for serving on my committee and providing helpful feedback. A special thanks to Denny Proffitt for teaching me the Gibsonian approach and supporting me in taking my crazy research ideas and helping me turn them into testable projects. To Jim Freeman, for taking me under his wing and providing a wonderful example of what a teacher should be. And finally, to Dan Willingham whose wisdom, advice, and witty jokes always helped to keep the important things in perspective.

I would also like to thank my parents for their unwavering support and never-ending love. It is your love and support that enabled me to keep dreaming and reaching for the stars. I can never repay you for everything you have given me, but I will spend the rest of my life trying. Thank you for not only being two of the best parents that one could ask for, but also two of the best friends one could ever have. I love you both.

I am also grateful and blessed to have such an amazing, supportive family.

Grandma, Grandpa, Aunt Karen, Kate, Sarah, my other mom (Carmen) and the entire Gaba clan – thank you for all your love and encouragement. I love you!

To Sally, Jon, Jeff, Max, Eric, Mollie, Allison, John, and Roberta – where would I have been without the late nights, barbecues, football tailgates, dancing, and vacations? A heartfelt thanks to my sister from another mister - Kathy. Thanks for the support babe – I finally made it to the other side!

To my friends and colleagues in Gilmer, I thank you for your help and support. A special thanks to Sally Linkenauer, Jon Bakdash, Jeff Pink, Janet Trammel, Lacy Krueger, Justin Storbeck, Jeanine Skorinko, Kate Ratliff, Colin Smith, Nicole Lindner, and Carlee Hawkins – you helped to make my journey a little bit easier.

To my wonderful research assistants who have helped make my research not only easier, but more fun. To Max Besse, Charles Koransky, Mike Kovacs, Mallory McFarland, Vicky Nguyen, and Evan Vitiello – you rock! Whether it was coding, scheduling, or running an experiment where you had to appear like you – ahem – just let one rip, I am thankful for your tireless efforts and hard work. I'd also like to thank Mike Hunter for his quantitative expertise.

To my kitties, Cassie, Ceebie, Charlie, and Ty whose fluffiness, cuddliness, and playful antics always make me smile. There's definitely truth in the quote of "Who rescued who?" We may have given you a home, but you have helped fill it with happiness and laughter. For all you have brought to us, I know the four of you surely rescued Jonathan and me.

And last, but certainly not least, to my husband, Jonathan. I dare say this dissertation would not have been possible without you by my side. I am deeply humbled by your love. You have given me strength when I had none, a shoulder to cry on when things got too tough, and laughter in those times when I needed it most. I am truly blessed to call you husband. You have not only made the past six years possible, but you have also made them worthwhile. I love you.

Chapter 1

In modern society, the normative value of being egalitarian is evident: schools launch tolerance programs with the goal of teaching younger individuals to go beyond another's skin color and gender (Flanagan & Faison, 2001), and many organizations and businesses routinely schedule diversity training events. Those who demonstrate bias are subject to consequences ranging from community reproach, job termination, or even legal recourse (Malos, 2007).

Despite this cultural value, discrimination still occurs. For example, Bertrand and Mullainathan (2004) showed that resumes from applicants with White sounding names (e.g., Emily) were 50% more likely to receive a call back from actual employers than the same resumes with Black sounding names (e.g., Lakisha). Similarly, Carpusor and Loges (2006) demonstrated that individuals with White sounding names were more likely to have landlords encourage them to visit a potential rental apartment than were applicants with stereotypically Black or Arab sounding names. These results suggest that individuals use stereotypes associated with Black people and Arabs to make decisions, which can result in discriminatory practices. The purpose of the present work is to investigate a new intervention aimed at changing beliefs and attitudes that may contribute to discrimination. In particular, this research examines whether educating people about how biased thinking can occur automatically or non-consciously can change their beliefs about bias and social judgments.

1.1 Egalitarian Values Do Not Guarantee Unbiased Thoughts or Behaviors

If people endorse egalitarianism, how could they still discriminate? One reason for the apparent discrepancy between individuals' values and overt behavior may be due,

in part, to the fact that both automatic and controlled processes influence judgment and decision-making (Devine, 1989; Epstein, 1994; Schneider & Schiffrin, 1977; Sloman, 1996).

1.1.1 Controlled Processes

Answering a question like “How much do you prefer White people to Black people?” requires introspection and the generation of a deliberate response. This intentional, deliberate responding reflects controlled processes. With controlled processes, the individual can choose which thoughts to express. How an individual chooses to respond indicates a conscious belief, or at least what the person is willing to report. Controlled processing explains only a portion of mental activity (Nisbett & Wilson, 1977; Nosek, 2007). People may not report all of the relevant mental activity in generating their explicit beliefs about White and Black people because they do not agree with some of it, or because they secretly agree but do not want others to know about those feelings, or because they are not even consciously aware of it.

1.1.2 Automatic Processes

In contrast to controlled processes, automatic processes characterize mental processes that occur without intention or control, and may require little to no conscious awareness to initiate or complete (Schneider & Schiffrin, 1977). Once a process becomes automatized, it is often difficult to suppress or modify (see, for example, Stroop, 1935). Importantly, information can be learned, encoded in memory, and become automatized even if the individual does not believe or want to use it (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). And, because this process is automatic, it may be

influential on behavior without the person's awareness, leading them to be mistaken about the reasons for their behavior (Nisbett & Wilson, 1977).

Thus, whereas controlled processes describe intentional and conscious thought, automatic processes describe thoughts and feelings that occur outside of conscious awareness or conscious control. For attitudes in particular, research has demonstrated that individuals' controlled attitudes (i.e., self-report) can be dissociated from their automatic ones (Cunningham, Preacher, & Banaji, 2001; Nosek, 2005; Nosek, Smyth, et al., 2007). Moreover, Greenwald, Poehlman, Uhlmann, and Banaji (2009) showed that across 32 studies involving socially sensitive contexts (e.g., race) automatic attitudes were *more* predictive of behavior and judgments than their controlled counterparts. One obvious interpretation of these findings is that individuals are concerned with appearing nonprejudiced and, as a result, deliberately respond in ways to make them appear egalitarian. However, another possibility is that because individuals are not always consciously aware of their mental processes, their automatic attitudes revealed may actually be counter to individuals' self-report. For example, individuals may observe a high proportion of criminals who are Black; in turn, individuals may be more likely to evaluate a Black person with an implicit assumption associating them with criminality even if they do not consciously believe it. Although individuals may consciously endorse egalitarian values, they may not be aware, or be able to inhibit the expression of associations learned through experience.

1.2 Education and Bias

Given that discrimination occurs despite individuals' reporting impartiality, many organizations have looked to diversity training programs as way to reduce discrimination. A recent survey shows that 71% of companies report engaging in some type of diversity training, with 68% of those requiring employee participation (SHRM, 2010). There are many approaches to diversity training, with some showing weak evidence of being effective. For example, multicultural education programs that focus on intergroup contact have shown some success at changing its participants' attitudes and behaviors (see Stephan, Renfro, & Stephan, 2004 for a review). However, for most organizations, diversity training is learning based, designed to reduce discrimination by having individuals replace existing preferences for some social groups over others with egalitarian attitudes, or at least motivate egalitarian behavior (see Jackson & Joshi, 2010 for a review). This type of diversity training presumes individuals are aware of their attitudes and the causes of their behavior and simply lack the motivation to change them (Kalev, Dobbin, & Kelly, 2006). The underlying assumption is that discriminatory practices are the result of intentional acts. As a result, most diversity training focuses on alteration of explicit, controlled prejudicial thoughts or behaviors. However, changing explicit biases may not be sufficient to eliminate discriminatory behavior. In fact, given the pervasiveness of egalitarian norms, many people are already explicitly egalitarian. As such, there are no explicit prejudices to remove. If people believe that explicit feelings and conscious intentions are necessary mechanisms for producing discriminatory behavior, then they may falsely believe that their behavior already reflects normative

values of non-discrimination. Therefore, diversity training that focuses exclusively on explicit beliefs and deliberate intentions may be ineffective because they fail to educate people on the automatic processes that can produce discriminatory behavior in people who are explicitly egalitarian and do not intend to discriminate.

In the current dissertation work, I am looking to develop and evaluate an approach that focuses on direct teaching/educating – a format that the existing literature does not find to be effective or evaluated. Only 36% of the organizations that provide diversity training report taking measures to assess whether the training was effective (SHRM, 2010). In a number of cases, *effectiveness* was measured by participants' self-assessment of whether they enjoyed the session, learned something, or felt that it was useful. For the small amount of research on effectiveness using actual behavioral outcomes such as changes in the diversity of the workplace, the evidence suggests that most existing diversity training practices are inert, with some even having counterproductive effects (Kalev, Dobbin, & Kelly, 2006; Naff & Kellough, 2003). Some evidence suggests that diversity training can *increase* individuals' biases toward minority group members (Kidder, Lankau, Chrobot-Mason, Mollica, & Friedman, 2004; Sidanius & Pratto, 1999). The root of the problem may be that diversity training is often a trainer's "best guesses" about practices that could reduce discrimination, rather than being grounded in research evidence of what changes discriminatory behavior (Kalev et al., 2006). In this dissertation, I will use scientific research to design a diversity program toward educating people about bias that is different from current existing practices. In addition, I will empirically evaluate its effectiveness.

1.2.1 The Bias Blindspot

Why does discrimination continue even after exposure to a diversity program?

One explanation for the discrepancy between values and behavior may be the “bias blind spot” (Pronin & Kugler, 2007), which suggests that individuals more easily recognize bias in others than they recognize in themselves. Indeed, many people believe that they are objective and impartial and, as a consequence, falsely conclude that they are immune to biased judgments based on social group memberships. For example, Uhlmann and Cohen (2007) showed that participants who affirmed their objectivity before evaluating a male and female job candidate were significantly *more* likely to hire the male candidate compared to the participants that were not asked to affirm their feelings of objectivity. Uhlmann and Cohen (2007) conclude that people who believe they are impartial adopt an “I think therefore it’s true” mindset (p. 218) whereby because they perceive themselves as objective, they believe that their thoughts are free from bias – even stereotypical ones – and consequently act on them. These results suggest that if diversity training increases people’s self-perceived objectivity, without actually changing *actual* objectivity, the training may actually have counterproductive effects. Individuals, not recognizing that they may have automatic stereotypes that influence judgment, may operate with the false belief that their intention to be objective is sufficient to reduce or eliminate discriminatory actions.

1.2.2 Experiential Education about Unintentional Bias may Change Beliefs about Bias

Given that most diversity programs focus on teaching individuals about intentional bias, the goal of the present work was to examine whether integrating material regarding automatic biases provided more behaviorally realistic training and education. Specifically, this dissertation investigated whether educating individuals about automatic bias changed participants' *beliefs about bias*, and in turn, their social behavior.

In contrast to typical diversity interventions that use non-expert “best guesses” to train participants, this education (a) is grounded in the latest science research showing how bias works, and (b) made it engaging and impactful by providing the lessons experientially rather than just describing the evidence. Previous findings have shown that individuals learn better when the material is given experientially rather than simply described (de Haan & de Ridder, 2006; DeRue & Wellman, 2009; Jarvis, 1987; Kolb, 1984; Pratt, 1993). As part of the bias education, I included material I believed would change four beliefs about bias. They included beliefs about acknowledgement that automatic bias exists, the recognition that automatic bias can influence behavior, self-perception of objectivity, and motivation to address automatic bias.

1.2.2.1 Acknowledgement that Automatic Bias Exists

Most diversity training focuses on intentional bias (Jackson & Joshi, 2010; Kalev, Kelly, & Dobbins, 2006; Von Bergen, Soper, & Foster, 2002). As a result, participants may leave “typical” training with a false belief that 1) bias is controllable and 2) as long as they value objectivity and fairness, discrimination will not occur. As demonstrated by

Uhlmann and Cohen (2007), this is not true; in fact, feelings of objectivity can lead to an increase in discrimination. One of the main themes of the current education is to teach individuals that bias can occur unintentionally. This is accomplished with descriptions of research evidence and direct experiential examples showing how this can occur – such as a group-delivered Implicit Association Test (IAT; Greenwald, et al., 1998). Thus, I expect that educated participants would be more likely to acknowledge that automatic bias exists.

1.2.2.2 Recognition that Automatic Bias can Influence Behavior

Previous research has shown that individuals believe themselves to be less susceptible to situational biases than their peers (Epley & Dunning, 2000; Pronin, Kugler, Savitsky, & Ross, 2001; Pronin et al., 2002). However, more recent research has demonstrated the value of educating individuals on these types of biases: individuals who learned about contextual effects were more likely to report that they could be susceptible to such effects compared to individuals who never received education (Pronin & Kugler, 2007). Given that part of the bias education demonstrates evidence showing how unintentional bias can influence individuals' judgments and behavior, I expect that this educational intervention would produce similar effects to Pronin and Kugler (2007), with educated individuals reporting stronger beliefs that bias can influence behavior compared to non-educated individuals.

1.2.2.3 Self-Perception of Objectivity

Many people believe that they are objective and immune to bias (Pronin & Kugler, 2007). By providing group demonstrations designed to show individuals that they

are not always aware of how their mind processes information, participants can learn about ways that they may unintentionally act in a biased manner. If the presentation presents the research evidence effectively and non-threateningly, then it would be difficult to escape the conclusion that “being objective” requires more than just intending to be objective. Thus, I expect that individuals who receive the scientifically-based bias education would be less likely to see themselves as objective compared to individuals who were not educated.

1.2.2.4 Motivation to Address Automatic Bias

Traditional forms of diversity training typically focus on historical prejudices, with the assumption being that White people and men are the problem (for race and gender respectively; for a review see Von Bergen, Soper, & Foster, 2002). As a result, this type of training may produce reactance and defensiveness among Whites and men. In contrast to these methods, the current education makes clear that bias is a human phenomenon. It does not single out any group as being susceptible to bias, and suggests that biased behavior is not just the product of a few “bad apples.” I suspect that this removes some of the negative stigma that only “bad people” or White people are biased. As a result, I expect that participants who receive automatic bias education would believe that even “good” people, including them, are susceptible to bias. This may change the minds of people who had not previously thought that they were biased and therefore did not feel motivated to address it. Recognizing a potential discrepancy in their egalitarian values and their potential to behave in a biased manner may increase their motivation to address bias (Monteith 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 2001). Thus,

they should be more motivated to address bias than individuals who did not receive the education.

1.3 Automatic Bias Education and Changing Beliefs about Bias May Change Social Judgment and Behavior

There are several reasons to hypothesize that automatic bias education would be more effective than traditional forms of diversity training in changing individuals' social judgments and behaviors. Wilson and Brekke (1994) argue that unwanted judgments and behaviors are difficult to avoid because, in part, individuals are not always aware of the processes by which these automatic responses occur. They provide a model suggesting that individuals can change these automatic responses through a series of steps. The first is to be aware of the potential for biased processing. Educating individuals about automatic bias directly addresses this step. In contrast to traditional forms of diversity training that assumes individuals know about their biases, this education provides participants with insight showing that they may have biases – biases that they may not be aware of or endorse.

Motivation to address the bias is step two of Wilson and Brekke's (1994) model. Monteith's (1993) self-regulation model and research provide evidence aligned with Wilson and Brekke's second step suggesting that, for some individuals, being aware of their biases motivates them to correct for it in an effort to reduce discriminatory effects on future behavior. For example, Monteith, Ashburn-Nardo, Voils, and Czopp (2002) showed, among individuals who value responding in nonprejudiced ways, evidence that they could be discriminatory created feelings of discord between how they think they

should and would behave. When faced with a subsequent bias-relevant decision, their motivation to be unbiased led these individuals to take more time in considering their potential behavioral responses. Taking extra time helped those individuals ensure that their outward response reflected their nonprejudiced internal values. Translating these findings to the proposed intervention, for some people, bias education may allow them to acknowledge the possibility that they harbor automatic biases that are in direct contradiction with their personal values (i.e., acknowledging that bias exists within themselves). By helping individuals become aware that they are not as objective as they may believe, they may become more motivated to address bias.

1.4 Why Education Alone Might Not Change Individuals' Social Judgment and Behavior

Despite my predictions that the educational intervention will address the first two steps of Wilson and Brekke's model (1994) for reducing the influence of biased thinking on judgments, it is possible that educating individuals about automatic biases will not be enough to change social judgment and behavior. Although bias education might make individuals aware of unwanted bias and motivate them to correct for it, Wilson and Brekke's model (1994) states that there are an additional two steps that must be met for individuals to be able to change their biased responses: they must be aware of the direction and magnitude of the bias, and be able to adjust their bias accordingly. Thus, even if the education is effective in addressing the first two steps, people may still not be able to detect the magnitude of the bias or know how to adjust their response to compensate for it. In courtroom research, for example, people find it difficult to disregard

material that they have been instructed to ignore (Kassin & Sommers, 1997; Pickel, 1995; Sue, Smith, & Caldwell, 1973). The idea is that the judge's instructions to disregard produce a "backfire effect" (Cox & Tanford, 1989) whereby individuals pay *more* attention to the evidence than if the judge had said nothing regarding the evidence. And, similar to Wilson and Brekke's (1994) fourth step in their mental contamination model, even when people agree that they should disregard information, they may not know how to "undo" the influence in their judgment.

So, it is possible that I will find evidence for change on beliefs and motivation, but not on social judgment. Even if this were the case, the intervention would still have promise as a first step in interventions that could ultimately reduce discrimination. After all, the first two steps of Wilson and Brekke's model must be achieved before the last two can be addressed.

Chapter 2

The purpose of the present work is to investigate whether education about the role of automatic processes in producing bias changes people's beliefs about bias, and subsequent decision-making. Study 1 and 2 were designed to examine whether educating individuals about automatic bias can change individuals' self-reported beliefs, attitudes, and decisions. Specifically, Study 1 tested whether such training is effective by comparing individuals' baseline beliefs regarding bias to individuals who report their beliefs after the intervention. Study 2 is a follow-up study assessing whether education has a lasting impact on individuals' beliefs and whether it can influence individuals' behavioral decisions. In addition, Study 2 will assess whether there are individual

differences in people's willingness, or motivations, to accept that automatic bias exists. Finally, Studies 3-6 experimentally examined the effectiveness of a shortened version of the intervention.

2.1 Study 1: Automatic Bias Education Influences People's *Beliefs about Bias*

The purpose of Study 1 was to investigate whether educating individuals about automatic bias could influence their attitudes and *beliefs about bias*, particularly their own susceptibility to bias. To test, incoming University of Virginia first year undergraduates attended a lecture with scientific background and demonstrations showing that their minds are prone to automatic biases. The presentation utilized visual and cognitive illusions highlighting how the mind processes information automatically. Then, it described how social biases operate similarly, and can occur even without awareness. To measure the short-term impact of the education, participants were randomly assigned to complete measures of their *beliefs about bias* within a few days before or after the presentation.

In addition, Study 1 examined the factor structure of the *beliefs about bias* questionnaire using exploratory and confirmatory factor analysis. This questionnaire was designed with four factors in mind: (1) *Acknowledgement that Automatic Bias Exists* (2) *Recognition that Automatic Bias can Influence Behavior* (3) *Self-Perception of Objectivity* and (4) *Motivation to Address Automatic Bias*. To see whether the questions related in ways I did not anticipate, an exploratory factor analysis was used on half of the responses for both the pre-test and post-test data collections. Then, to confirm the

appropriateness of that factor structure, a confirmatory factor analysis was applied to the other half of the pre- and post-test responses.

The actual factor structure turned out slightly differently than hypothesized. Therefore, to anticipate those results, the factors that are featured in the results of this section and for the rest of the paper are: (1) *I Have Bias*, (2) *Automatic Bias Influences Me*, (3) *Self-Perception of Objectivity*, (4) *Motivation to Address Automatic Bias*, and (5) *Knowing and Preventing Automatic Bias*. The first, second, and fifth of these are three components that I had conceptualized as being just two components initially – *Acknowledgement that Automatic Bias Exists* and *Recognition that Automatic Bias can Influence Behavior*. In addition, there were two items that did not load onto a factor but were still of interest. These items were examined independently: “A good person is someone who does not have bias” and “I am comfortable talking about potential biases I may have with other people.”¹

2.1.1 Method

Participants

Potential participants were the entire first year entering class of undergraduates from the University of Virginia in the Fall of 2010. Part of orientation program was a lecture titled “Mindbugs” delivered by Brian Nosek to approximately 3,500 people, most of whom were the first-year students.

There were a total of 3,246 first year students in the class of 2014. Most of those, what precise percentage I cannot be sure, attended this educational lecture. Of the

---

¹ While I continued to collect data on these two items during Studies 2-6, my analysis focused on examining factor differences. Thus, these items were not examined independently. Rather, these items were included in participants’ aggregate BaB score.

attendees, a total of 2,194 surveys were returned (68% of total possible sample). Whether an individual survey was administered and returned was mostly influenced by the Resident Advisor (RA) that was responsible for administering the survey to his or her orientation group. This reduces individual self-selection effects in whether to complete the survey or not. Details of survey distribution and sampling appear in the Procedure section below.

The sample is not randomly selected from the population, but it does have some advantages over typical laboratory samples. The primary selection effects were whether the student was admitted to UVa and chose to attend in the Fall of 2010. This contrasts from most psychology samples that are recruited from currently enrolled students in a psychology class. Those samples (1) self-select into the class presumably from interest (e.g., people who choose to pursue psychology may be more likely to be female and liberal compared to the average student); and, (2) learn, through participation in their psychology course, information related to cognitive and social biases – such as the foundations of stereotyping and prejudice. These selection biases become more important in Study 2.

Of the participants (945 male, 2 unreported), 252 (11.5%) reported being East Asian, 114 (5.2%) reported being South Asian, 139 (6.3%) reported being Black or African American, 1,450 (66.1%) reported being White, 18 (1.0%) being biracial – Black/White, 117 (5.3%) reported being more than one race – other, 48 (2.2%) reported not knowing, and 56 (2.6%) chose not to respond. 115 (5.2%) participants reporting being of Hispanic or Latin descent, 1,634 (74.5%) participants reported not being of

Hispanic or Latin descent, 93 (4.2%) participants reported not knowing and 362 (16.5%) chose not to respond. In addition, participants indicated if they were willing to be contacted for a follow-up study. Of the sample, 1,796 (81.9%) agreed to be contacted.

Materials

Bias Education. The education actively engages the audience with interactive demonstrations of how the mind uses automatic processes. The presentation includes three sections. The first section showed individuals how the mind integrates previous experience and the present context to make sense of what is going on in the environment. To illustrate this, cognitive and visual illusions were employed to demonstrate that mental experience is not the same as reality. For example, participants experienced the McGurk effect (McGurk & MacDonald, 1976) as well as Shepherd's (1990) table illusion, highlighting how the mind imposes interpretations on sensory information to construct its perceptual experience. During the second part of the presentation, individuals learned how automatic processes can lead to social bias. The key part of the second section is a group demonstration of the Implicit Association Test (IAT; Greenwald, et al. 1998). The IAT measures associations between concepts and evaluations without requiring an act of introspection about one's feelings. It can reveal associations that are quite different than self-reported attitudes and beliefs. In this presentation, the participants found it easier to categorize images of the University of Virginia with pleasant words compared to categorizing the pleasant words with their rival, Virginia Tech. This version of the IAT provides the participants with an understanding of how the task works. Then, the participants completed a second group

IAT that measured associations of pleasant words with Black and White people. As a group, the participants were able to categorize pleasant words with White people more quickly than pleasant words with Black people. In the final part of the presentation, individuals were shown how automatic bias influences decisions unintentionally with illustrative examples from scientific research.

Bias Questionnaire. Participants rated their agreement with questions using a 6-point Likert scale (1=strongly disagree, 6=strongly agree). The questionnaire assessed their attitudes and beliefs about the pervasiveness of bias. The questions assessed different aspects of beliefs about bias such as: acknowledgement that automatic bias exists (“Everyone, including me, has biases toward other people.”), recognition that automatic bias can influence behavior (“Unwanted biases may affect the way I make decisions without my realizing it.”), self-perception of objectivity (e.g., “I am an objective decision-maker”), and motivation to address automatic bias (“I am motivated to ensure that unwanted biases do not influence my decision making.”). Finally, all participants reported whether they were willing to be contacted with additional opportunities to learn about bias. This question was intended as a behavioral intention – a desire to learn more about automatic bias. For a complete list of questions, refer to Appendix A.

Demographics. Participants reported their race, ethnicity, and sex. In addition, participants indicated if they were willing to be contacted for a follow-up study.

Procedure

During the University of Virginia orientation, RAs coordinate activities for first-year students. There are about 170 RAs and a total of about 3,250 first-year students. Importantly, RAs were not aware of the study hypotheses, and were not directly involved in the research with the exception of distributing and returning the surveys in the assigned time frame. The materials were distributed to the RAs of each first year housing unit. All materials were condensed to a single sheet of paper with instructions and demographics on one side of the page and the bias questionnaire and behavioral measure on the reverse. RAs received an envelope with the surveys and instructions on the cover telling them when they should distribute it and how to return the materials. Half the envelopes instructed the RA to distribute the surveys sometime before the automatic bias education, the other half after the education. In addition, RAs noted when they distributed the questionnaire, how many surveys were distributed, and how many were collected.

Therefore, random assignment was at the group level, rather than by individual. There were approximately 170 groups, and I received survey packets back from 134 (79%, 65 pre-test groups and 69 post-test groups). Of these groups, the RAs who provided the total number of surveys they distributed, reported distributing a total of 2,111 surveys, and submitted 1,907 (90% total response rate of the reported distributed surveys). This design also reduces potential demand effects as the participants did not complete the measures in the presence of the automatic bias lecturer or other experimenters involved in the research. Participants completed the measures in the context of their orientation group, not a psychological laboratory. Pre-test surveys were

administered an average of 37.03 hours ($SD = 14.98$ hours, Range 0.75 – 122 hours) prior to the bias education. Post-test surveys were administered an average of 5.72 hours ($SD = 4.96$ hours, Range 2 – 46.25 hours) after the education.

2.1.2 Results and Discussion

Results

Ruling out Group-Level Effects

The intended research design is focused on individual effects of receiving the treatment, but functionally the data collection is nested in a way that could impact the results. Each participant was a member of an orientation group, and the leader of that orientation group administered the questionnaire. It is possible participants may be influenced by their group on our measures and be more like their own group members² than they are like members of other groups. For example, one member of a post-test group could agree with the educational message; as a result, he might positively convey his enthusiasm, convincing other group members it was a worthwhile presentation. To ensure I did not unintentionally attribute group effects to the individual, I first created a composite index of the *beliefs about bias* questionnaire by averaging all of the items after coding them in the direction of the hypothesized changes. With that score, I calculated an intraclass correlation to determine if group members are more similar to each other than to other individuals for the overall sample. The magnitude of the overall correlation was small, 0.13, suggesting little group-level variation. I also calculated the intraclass correlation for pre and post-test participants separately because group-level effects based

---

² Participants' group membership is determined by the resident assistant to which they were randomly assigned.

on my intervention would presumably be more likely to occur in the post-test groups. These also revealed small correlations, 0.05 and 0.10, respectively. Given these small correlations, I focused the rest of my analyses at the individual-level.

Did the Education Affect Individuals' Beliefs about Bias?

To assess whether the bias education changed individuals' *beliefs about bias*, I computed an overall aggregate score for all 18 questions to calculate individuals' overall change in their *Beliefs about Bias* (BaB) score where higher scores indicated greater change in the direction hypothesized *a priori* for each item. Results revealed that participants who completed the questionnaire after the education were more likely to have a higher BaB score ($M = 3.88$, $SD = 0.52$) than participants who completed the bias questionnaire beforehand ($M = 3.60$, $SD = 0.51$), $t(2192) = 13.01$, $p < 0.001$, $d = 0.56$. As shown on Table 1 below, item by item analyses show that 16 of the 18 items significantly changed in the hypothesized direction, with a Cohen's d ranging from 0.10 to 0.44 (average $d = 0.30$). No items changed in the direction opposite of my hypotheses.

Table 1. The Impact of Automatic Bias Education on Participants' Beliefs about Bias

	Before	After	Effect Size
My decision making is rational and objective.	5.29	5.10	0.21*
When I make decisions, I am aware of the reasons for those decisions.	5.43	5.24	0.24*
If my decision was influenced by unwanted biases, I would know about it.	3.89	3.39	0.40*
I am comfortable talking about potential biases I may have with other people.	4.32	4.44	0.10*
The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.	5.23	5.03	0.18*
Awareness of one's own unwanted biases is an important step for reducing their influence on behavior.	5.17	5.21	0.04
I am motivated to ensure that unwanted biases do not influence my decision making.	5.00	4.99	0.00
I am concerned that I may have unwanted biases that I am not aware of.	3.33	3.79	0.33*
Everyone, including me, has biases toward other people.	4.56	4.89	0.26*
Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.	4.81	4.98	0.18*
I am biased.	3.72	4.06	0.23*
Whether I want to or not, I have more positive feelings about some racial groups than others.	3.62	4.20	0.44*
A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.	4.62	4.81	0.18*
Whether I am aware of it or not, I use a person's race or ethnicity to form an impression of the kind of person they are.	3.07	3.56	0.37*
Unwanted biases probably influence my decision making about other people.	3.78	4.21	0.34*
If people try hard enough, their decisions about other people will not be influenced by unwanted biases.	4.19	3.89	0.23*
Unwanted biases may affect the way I make decisions without my realizing it.	4.19	4.56	0.34*
A "good person" is someone who does not have biases.	3.40	3.08	0.23*
Percent agreeing to be contacted to participate in additional bias research	55	57	0.41

*significant at the $p < 0.05$ level

Are there Factors in the Items that Provide a Better Summary of Individuals' Change in Their Beliefs about Bias?

The item-by-item analysis suggests that there was variation in the extent to which different beliefs changed in response to the education. Rather than evaluate each item individually, I next tested whether there were sensible latent factors that aggregated items to provide a simpler representation of people's *beliefs about bias*. Although the questionnaire was designed to represent four related, but distinct factors, I first conducted an exploratory factor analysis (EFA) to determine whether there was an alternative structure that was not anticipated. Because the intervention could have influenced the factor structure, I conducted this analysis separately for participants who completed the questionnaire before the bias education and participants who completed the questionnaire afterwards. Also, to strengthen confidence in the appropriateness of the factor structure, I divided both the pre-test and post-test groups in half. I conducted the EFA on the first half of each sample, and then conducted a confirmatory factor analysis (CFA) with the identified factors on the second half of the sample. The 18 items were analyzed with a principal axis factoring method in SPSS using a Promax rotation. This rotation allowed the factors to be correlated. In line with Kaiser's criterion (1960), all factors with eigenvalues greater than 1.0 were retained.³

Exploratory Factor Analysis with "Before" Participants. As shown on Table 2 below, the initial analysis on half of the "before" participants revealed five factors with eigenvalues over 1. The first factor accounted for 23.11% of the variance (eigenvalue of

---

³ Although there has been debate about which factor extraction method is best, Kaiser's criterion is considered to be an accurate measurement technique when it (a) includes a few factors to the structure and (b) tests a large sample (for a review see Hill & Lewicki, 2006).

4.16). The second factor accounted for 13.16% of the variance (eigenvalue of 2.37). The third factor accounted for 8.16% of the variance (eigenvalue of 1.47). The fourth (eigenvalue of 1.27) and fifth factor (eigenvalue of 1.18) accounted for 7.05% and 6.53% of the variance respectively. The item “I am comfortable talking about potential biases I have” was dropped from the factor structure since it failed to load onto any of the factors with a factor loading of 0.30 or above. In addition, the item “A ‘good’ person is someone who does not have biases” was dropped as it cross loaded onto three factors. Finally, the items “Whether I want to or not, I have more positive feelings about some racial groups than others” and “Unwanted biases probably influence my decision making about other people” cross loaded onto factors one and two. In this case, their placement was determined by which factor it loaded more strongly.

Table 2. Factor Loadings for the Exploratory Factor Analysis of “Before” Participants

	Factor Loading
Factor 1	
Whether I want to or not, I have more or not, I have more positive feelings about some racial groups than others.	.493
I am biased.	.818
Everyone, including me, has biases toward other people.	.771
Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.	.376
Factor 2	
A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.	.441
Unwanted biases probably influence my decisions about other people.	.453
Unwanted biases may affect the way I make decisions without realizing it.	.752
Whether I am aware of it or not, I use a person’s race or ethnicity to form an impression of the kind of person they are.	.441
I am concerned that I have unwanted biases that I am not aware of.	.637
Factor 3	
The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.	.443
I am motivated to ensure that unwanted biases do not influence my decision making.	.699
Awareness of one’s own unwanted biases is an important step for reducing their influence on behavior.	.833
Factor 4	
My decision making is rational and objective.	.677
When I make decisions, I am aware of the reasons for those decisions.	.729
Factor 5	
If people try hard enough, their decisions about other people will not be influenced by unwanted biases.	.614
If my decision was influenced by unwanted biases, I would know about it.	.478

Exploratory Factor Analysis with “After” Participants. The initial analysis with half of the “after” participants also revealed five factors with eigenvalues over 1. The first factor accounted for 22.98% of the variance (eigenvalue of 4.14). The second factor accounted for 13.76% of the variance (eigenvalue of 2.48). The third factor accounted for 8.41% of the variance (eigenvalue of 1.51). The fourth (eigenvalue of 1.20) and fifth factor (eigenvalue of 1.16) accounted for 6.64% and 6.44% of the variance respectively.

This analysis revealed a very similar structure as was observed with the “before” participants. For example, the item “I am comfortable talking about potential biases I have” was dropped from the factor structure since it failed to load onto any of the factors with a factor loading of 0.30 or above. Likewise, the item “A ‘good’ person is someone who does not have biases” was dropped as it cross loaded onto three factors. Finally, the items “Whether I want to or not, I have more positive feelings about some racial groups than others” and “Unwanted biases probably influence my decision making about other people” cross loaded onto factors one and two. In this case, its placement was determined by which factor it loaded more strongly. For a complete description of the loadings and their factors, refer to Table 3 below.

Table 3. Factor Loadings for the Exploratory Factor Analysis of “After” Participants

	Factor Loading
Factor 1	
A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.	.482
Unwanted biases probably influence my decisions about other people.	.528
Unwanted biases may affect the way I make decisions without realizing it.	.748
Whether I am aware of it or not, I use a person’s race or ethnicity to form an impression of the kind of person they are.	.396
I am concerned that I have unwanted biases that I am not aware of.	.580
Factor 2	
Whether I want to or not, I have more or not, I have more positive feelings about some racial groups than others.	.433
I am biased.	.811
Everyone, including me, has biases toward other people.	.617
Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.	.331
Factor 3	
The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.	.514
I am motivated to ensure that unwanted biases do not influence my decision making.	.763
Awareness of one’s own unwanted biases is an important step for reducing their influence on behavior.	.636
Factor 4	
My decision making is rational and objective.	.746
When I make decisions, I am aware of the reasons for those decisions.	.726
Factor 5	
If people try hard enough, their decisions about other people will not be influenced by unwanted biases.	.401
If my decision was influenced by unwanted biases, I would know about it.	.547

Interpretation. After conducting exploratory factory analyses for participants who completed the bias questionnaire before and after the education separately, results

revealed that the same five factors emerged for both groups. Before conducting the confirmatory factor analyses on the second half of participants from each group, I reviewed the factors to see if they had interpretable meaning.

The first factor of five items included: “A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of”, “Unwanted biases probably influence my decision making about other people”, “Unwanted biases may affect the way I make decisions without realizing it”, “Whether I am aware of it or not, I use a person’s race or ethnicity to form an impression of the kind of person they are”, and “I am concerned that I may have unwanted biases that I am not aware of.” These five questions are similar in that they focus on unwanted biases and whether individuals believe they can be influenced by them. Therefore, I named this first factor, *Automatic Bias Influences Me*.

The second factor was comprised of four items: “Whether I want to or not, I have more or not, I have more positive feelings about some racial groups than others”, “I am biased”, “Everyone, including me, has biases toward other people”, and “Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.” These questions targeted participants’ beliefs that they were biased. Thus, I named this factor *I Have Bias*.

The third factor included three items: “The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly”, “I am motivated to ensure that unwanted biases do not influence my decision making”, and “Awareness of one’s own unwanted biases is an important step for

reducing their influence on behavior.” All items referred to elements of motivations for dealing with bias, so I named this factor *Motivation to Address Automatic Bias*.

The fourth factor included two items: “My decision making is rational and objective” and “When I make decisions, I am aware of the reasons for those decisions.” This factor was easily identified as *Self-Perception of Objectivity*.

Finally, the fifth factor included two items: “If people try hard enough, their decisions about other people will not be influenced by unwanted biases” and “If my decision was influenced by unwanted biases, I would know about it.” These questions assessed knowledge about bias influence and whether it can be prevented. As such, I named the factor *Knowing and Preventing Automatic Bias*.

These factors were close, but not perfectly aligned with the four factors that I had hypothesized originally. The biggest differences were the presence of a fifth factor making a distinction between the original factors *Acknowledgement That Automatic Bias Exists* and *Recognition that Automatic Bias can Influence Behavior*. Instead of these two factors, the EFAs suggest that these factors are better represented as three: *I Have Bias*, *Automatic Bias Influences Me*, and *Knowing and Preventing Automatic Bias*.

Confirming the Five Factor Structure with Confirmatory Factor Analysis

After conducting the exploratory factor analysis on half of each of the “before” and “after” samples, I wanted to ensure that the proposed factor structure provided a good model to the remainder of the data. Confirmatory factor analysis allowed me to statistically test the fit of the theoretical model proposed by my exploratory analysis against my actual data. Thus I pursued a confirmatory factor analysis on the remaining

half of both subsamples – participants who completed the questionnaire before the bias education and participants who completed the questionnaire afterwards – separately.

To assess the goodness of fit of the five-factor solution, the 16 items that had loaded onto a factor in the EFAs were submitted to confirmatory factor analytic procedures using Proc Calis in SAS representing the model identified by the exploratory factor analysis. The goodness of fit of the model to the data was evaluated using a goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI). These measures are standardized and range from 0 to 1, with 1 indicating a perfect fit. GFI and AGFI values of 0.85 and above are considered to reflect a good fit of the model to the data.

This initial model fit the data well for “before” participants (GFI = 0.90, AGFI = 0.86) and “after” participants (GFI = 0.91, AGFI = 0.87), suggesting that the five-factor structure was an appropriate model to apply to the data to test how the individuals’ beliefs about bias changed as a function of the automatic bias education they received. See Table 4 and 5 for a complete list of the factor loadings and residuals revealed by the CFA for “before” and “after” participants, respectively.

Table 4. Factor Loadings for the Confirmatory Factor Analysis of “Before” Participants

	Factor Loading	Residual Error
Automatic Bias Influences Me		
A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.	0.32	0.95
Unwanted biases probably influence my decision making about other people	0.82	0.57
Unwanted biases may affect the way I make decisions without realizing it.	0.74	0.67
Whether I am aware of it or not, I use a person’s race or ethnicity to form an impression of the kind of person they are.	0.61	0.79
I am concerned that I may have unwanted biases that I am not aware of.	0.44	0.90
I Have Bias		
Whether I want to or not, I have more or not, I have more positive feelings about some racial groups than others.	0.62	0.79
I am biased.	0.69	0.72
Everyone, including me, has biases toward other people.	0.70	0.71
Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.	0.57	0.82
Motivation to Address Automatic Bias		
The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.	0.63	0.77
I am motivated to ensure that unwanted biases do not influence my decision making.	0.70	0.71
Awareness of one’s own unwanted biases is an important step for reducing their influence on behavior.	0.72	0.70
Self-Perception of Objectivity		
My decision making is rational and objective.	0.57	0.82
When I make decisions, I am aware of the reasons for those decisions.	0.90	0.44
Knowing and Preventing Automatic Biases		
If people try hard enough, their decisions about other people will not be influenced by unwanted biases.	0.50	0.87
If my decision was influenced by unwanted biases, I would know about it.	0.61	0.79

Table 5. Factor Loadings for the Confirmatory Factor Analysis of “After” Participants

	Factor Loading	Residual Error
Automatic Bias Influences Me		
A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.	0.30	0.95
Unwanted biases probably influence my decision making about other people	0.77	0.64
Unwanted biases may affect the way I make decisions without realizing it.	0.75	0.66
Whether I am aware of it or not, I use a person’s race or ethnicity to form an impression of the kind of person they are.	0.64	0.77
I am concerned that I may have unwanted biases that I am not aware of.	0.42	0.91
I Have Bias		
Whether I want to or not, I have more or not, I have more positive feelings about some racial groups than others.	0.64	0.77
I am biased.	0.72	0.69
Everyone, including me, has biases toward other people.	0.70	0.71
Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.	0.61	0.80
Motivation to Address Automatic Bias		
The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.	0.44	0.90
I am motivated to ensure that unwanted biases do not influence my decision making.	0.78	0.63
Awareness of one’s own unwanted biases is an important step for reducing their influence on behavior.	0.70	0.72
Self-Perception of Objectivity		
My decision making is rational and objective.	0.96	0.27
When I make decisions, I am aware of the reasons for those decisions.	0.43	0.87
Knowing and Preventing Automatic Biases		
If people try hard enough, their decisions about other people will not be influenced by unwanted biases.	0.68	0.74
If my decision was influenced by unwanted biases, I would know about it.	0.57	0.82

Correlations between Scale Items

After confirming my five-factor structure, I computed an average score for the items representing each factor. The correlations between the factors are shown on Table 6. The factors *I Have Bias* and *Automatic Bias Influences Me* showed strong positive correlations, suggesting that they measured similar beliefs. The factors *I Have Bias* and *Automatic Bias Influences Me* were negatively related to *Knowing and Preventing Automatic Bias*. This suggests that the more an individual believed they had bias or were susceptible to automatic bias, the less likely they believed they would be aware of its effects. *Motivation to Address Automatic Bias* and *Self-Perception of Objectivity* showed weak relations with all the other factors. This provides evidence that motivation to address bias is independent of the extent to which people believe that it exists or that they are biased. Further, it is notable that self-perceptions of objectivity are unrelated to beliefs that one has automatic bias. Intuitively, these would seem to be opposing concepts, but the lack of relationship suggests that these beliefs are independent. This observation is relevant for evidence in Study 2 showing that change from the intervention was durable for believing that one had automatic bias, but not for self-perceptions of objectivity.

Treating the beliefs as a single 18-item scale revealed moderately good internal consistency ($\alpha = 0.74$). While the factor structure suggests that it is not a single factor, I do analyze and report the aggregated measure as an “overall” analysis of change in beliefs about bias before analyzing each factor independently. The results showed fair to good internal consistency for each of the factors individually except for one (*I Have*

Bias, $\alpha = 0.74$, *Automatic Bias Influences Me*, $\alpha = 0.74$, *Self-Perception of Objectivity*, $\alpha = 0.68$, *Motivation to Address Automatic Bias*, $\alpha = -0.21^4$, *Knowing and Preventing Automatic Bias*, $\alpha = 0.53$).

Table 6. Correlations Between Factors

	I Have Bias	Automatic Bias Influences Me	Self-Perception of Objectivity	Motivation to Address Automatic Bias	Knowing and Preventing Automatic Bias
I Have Bias	1	0.63 ⁺	-0.05*	0.03	-0.29 ⁺
Automatic Bias Influences Me		1	-0.12 ⁺	0.13 ⁺	-0.33 ⁺
Self-Perception of Objectivity			1	0.06**	0.17 ⁺
Motivation to Address Automatic Bias				1	-0.01
Knowing and Preventing Automatic Bias					1

*significant at $p \leq 0.05$

**significant at $p \leq 0.01$

⁺significant at $p \leq 0.001$

Each of the Beliefs about Bias Changed Following the Intervention

I conducted *t*-tests on each factor to discern whether there were differences between participants who completed the bias questions before the education compared to participants who completed the questionnaire afterwards.

Participants completing the questionnaire after the education more strongly agreed that *I Have Bias* ($M = 4.53$, $SD = 0.91$) than participants who completed the

⁴ When the variables are coded such that higher values mean greater change in the predicted direction from the intervention (confirmed in Study 1), one of the items (“The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly”) is negatively related to the other two (“I am motivated to ensure that unwanted biases do not influence my decision making”, “Awareness of one’s own unwanted biases is an important step for reducing their influence on behavior”). Although they loaded together in the factor analysis because of their strong relations, the fact that they change in the same direction despite being negatively correlated creates an odd scenario for typical calculation of alpha. When they are coded so that they are all positively related, the alpha is 0.68; however, when coded so that positivity indicates the direction of predicted and observed change following the intervention, the alpha is -0.21.

questionnaire beforehand, ($M = 4.18$, $SD = 0.97$), $t(2121) = 8.69$, $p < 0.001$, $d = 0.38$.

Likewise, participants who completed the questionnaire after the education more strongly agreed that *Automatic Bias Influences Me* ($M = 4.18$, $SD = 0.83$) than participants who completed the questionnaire beforehand, ($M = 3.80$, $SD = 0.90$), $t(2100) = 10.45$, $p < 0.001$, $d = 0.46$). Results also revealed that participants completing the questionnaire after the education showed more agreement in *Motivation to Address Automatic Bias* ($M = 4.05$, $SD = 0.58$) than participants completing it beforehand ($M = 3.98$, $SD = 0.51$), $t(2191) = 3.21$, $p < 0.01$, $d = 0.14$.

Similarly, participants who completed the bias questionnaire after receiving education showed less agreement in *Self-Perception of Objectivity* ($M = 5.17$, $SD = 0.79$) than participants who completed the questionnaire beforehand, ($M = 5.36$, $SD = 0.68$), $t(2188) = 5.94$, $p < 0.001$, $d = 0.25$. Finally, “after” participants showed less agreement in *Knowing and Preventing Automatic Bias* ($M = 3.64$, $SD = 1.09$) than did “before” participants, ($M = 4.04$, $SD = 1.02$), $t(2186) = 8.89$, $p < 0.001$, $d = 0.38$. All of the beliefs about bias changed, and most with effect sizes that were moderately sized (Cohen, 1988). The exception was *Motivation to Address Automatic Bias*, which changed, but only weakly ($d = 0.14$).

Does the Amount of Time Between Automatic Bias Education and Reporting Beliefs about Bias Matter?

One question is whether individuals’ *beliefs about bias* differed as a function of the amount of time that had passed from receiving the education and completing the bias questionnaire. This question is a first step in assessing whether automatic bias education

has longer-term durability on participants' *beliefs about bias*. The majority of the sample completed the bias questionnaire the same day as receiving the education ($M = 5.72$ hours, $SD = 4.96$, Range 2 – 46.25 hours); as such, this question examines whether the intervention lasted a matter of hours. Results revealed that for participants who completed the questionnaire after receiving the bias education, the time between receiving the education and completing the bias questionnaire was not correlated to their beliefs *I Have Bias* ($r = 0.03$, $p = 0.27$), *Automatic Bias Influences Me* ($r = 0.00$, $p = 0.88$), *Self-Perception of Objectivity*, ($r = -0.02$, $p = 0.51$), *Motivation to Address Automatic Bias*, ($r = 0.05$, $p = 0.12$), or *Knowing and Preventing Automatic Bias*, ($r = -0.03$, $p = 0.39$). This suggests that having several hours between the automatic bias education and completion of the questionnaire did not influence participants' responses.⁵

It is important to note that the experimental conditions are confounded with time. That is, the control group that did not see the intervention was not measured at the same point in time as “after” participants. Thus, participants could already have experiences (e.g., meeting new people, diversity training unrelated to the intervention) that change their *beliefs about bias*. However, participants in the “before” and “after” conditions did not all complete the survey at the same time, providing an opportunity to see if there was change occurring in beliefs about bias during this short time period. The correlations with time within the “before” and “after” participants provide evidence that history effects are not a plausible counter explanation. The correlation between beliefs about bias and time

---

⁵ For participants who completed the questionnaire before receiving the bias education, the time between receiving the education and completing the bias questionnaire was significantly correlated for some, but not all, of the factors. Results revealed very small negative correlations for *I Have Bias* ($r = -0.11$, $p = 0.002$) and *Automatic Bias Influences Me* ($r = -0.11$, $p = 0.001$). Results also showed a small positive correlation on participants' beliefs about *Knowing and Preventing Automatic Bias* ($r = 0.09$, $p = 0.01$).

for “before” participants did show slight evidence for change on some factors, but it was very weak (all r 's ≤ -0.11). The results revealed no significant correlation of time for the “after” participants. As I will show in the next study, there is some evidence that beliefs about bias are changing over time, but the data shows that the time difference between pre and post-tests was not long enough for history effects to be a counter-explanation for the observed effects of the intervention.

Does Automatic Bias Education Influence Some Groups of People More than Others?

Another question worth assessing is whether automatic bias education affected participants differently. That is, are particular social groups (e.g., race, gender) more likely to show changes in their *beliefs about bias* compared to others? I did not have particular hypotheses about this in advance. I first tested whether gender or race moderates the aggregate *beliefs about bias* (BaB) score – as main effects and interactions with the intervention. Then, using the same type of analysis, I individually tested each factor. To reduce the likelihood of reporting spurious effects, I set a more conservative significance criterion ($\alpha = 0.01$). I report results for the participants' overall BaB score and significant findings below.

Gender. A 2 (Gender: Male versus Female) X 2 (Condition: Before versus After) ANOVA revealed no significant main effect of gender, $F(1, 2184) = 3.01, p = 0.08, d = 0.07$, or interaction of Gender x Condition, $F(1, 2184) = 1.84, p = 0.18, d = 0.06$, on participants' BaB score. Results did show several main effects of gender such that females showed higher agreement than males on *Automatic Bias Influences Me*, $F(1, 2184) = 8.49, p = 0.004, d = 0.12$, and *Motivation to Address Bias*, $F(1, 2183) = 22.42, p$

< 0.001 , $d = 0.20$. Females reported less agreement on *Self-Perception of Objectivity* $F(1, 2183) = 10.42$, $p = 0.001$, $d = 0.14$ than males, indicating they believed themselves to be less objective than male participants. Interestingly, results revealed a marginally significant Gender X Condition interaction on *Automatic Bias Influences Me*, $F(1, 2184) = 4.62$, $p = 0.03$, $d = 0.09$, indicating that females showed greater change than males in response to the intervention (pre-test females: $M = 3.81$, $SD = 0.89$, pre-test males: $M = 3.78$, $SD = 0.93$; post-test females: $M = 4.27$, $SD = 0.81$, post-test males: $M = 4.08$, $SD = 0.84$). It appears that females were more convinced by the presentation that their behavior is influenced by automatic biases. Likewise, results showed a marginally Gender X Condition interaction suggesting that females *Self-Perception of Objectivity* declined more in response to the presentation than males did: $F(1, 2183) = 5.44$, $p = 0.02$, $d = 0.10$, (pre-test females: $M = 5.35$, $SD = 0.65$, pre-test males: $M = 5.38$, $SD = 0.71$; post-test females: $M = 5.09$, $SD = 0.82$, post-test males: $M = 5.27$, $SD = 0.75$). Because I conducted multiple analyses that were not theory guided, these interactions should be interpreted with caution.

Race. A 7 (Race: East Asian versus South Asian versus Black versus White versus Biracial-Black/White versus More than One Race-Other versus Other/Unknown) X 2 (Condition: Before versus After) ANOVA revealed a significant main effect of race, $F(6, 2124) = 5.81$, $p < 0.0001$, $d = 0.10$, but no significant interaction of Race X Condition, $F(6, 2124) = 1.24$, $p = 0.20$, $d = 0.05$, on participants' BaB score. Overall, East Asian participants showed the highest BaB score ($M = 3.46$, $SD = 0.48$), while Black participants showed the lowest overall BaB score ($M = 3.28$, $SD = 0.52$).

In addition to participants' overall BaB score, results showed several small, but reliable main effects of race on *I Have Bias*, $F(6, 2124) = 6.03, p < 0.0001, d = 0.11$, *Automatic Bias Influences Me*, $F(6, 2124) = 4.81, p < 0.0001, d = 0.10$, and *Knowing and Preventing Bias*, $F(6, 2123) = 7.21, p < 0.0001, d = 0.12$.

Overall, results indicated that participants who identified as East Asian reported the highest levels of agreement on *I Have Bias* ($M = 4.50, SD = 0.88$) and *Automatic Bias Influences Me* ($M = 4.15, SD = 0.79$). Black participants showed the highest level of agreement on *Knowing and Preventing Automatic Bias* ($M = 4.15, SD = 1.12$). In contrast, Black participants showed the lowest levels of agreement on *I Have Bias* ($M = 4.04, SD = 1.05$). Participants who reported being part Black and part White showed the lowest level of agreement on *Automatic Bias Influences Me* ($M = 3.64, SD = 0.85$) while White participants reported the lowest level of agreement on *Knowing and Preventing Automatic Bias* ($M = 3.74, SD = 1.06$).

Although results revealed no significant interaction of Race X Condition, one group stood out as being particularly sensitive to the intervention. Specifically, Black participants, compared to the rest of the sample, showed the greatest change from pre to post on *I Have Bias* ($d = 0.61$ versus $d = 0.38$), *Automatic Bias Influences Me* ($d = 0.50$ versus $d = 0.46$), and *Self-Perception of Objectivity* ($d = 0.35$ versus $d = 0.25$) than any other race. Notably, the question “Whether I want to or not, I have more positive feelings about some racial groups than others” that was included in *I Have Bias* showed the greatest pre to post change ($d = 0.60$) compared to any other question that comprised this factor.

Discussion

In Study 1, I investigated whether educating individuals about automatic bias change their *beliefs about bias*. The education was successful: individuals completing the bias questionnaire after receiving the education were more likely to report that bias might influence them, believe that they had bias, and report more motivation to address bias than participants who completed it beforehand. Likewise, participants who completed the bias questionnaire after receiving education were less likely to believe they were objective and that they would know and be able to prevent bias than participants who completed the questionnaire beforehand.

The Effect of Education Impacted Individuals' Beliefs about Bias Differently

The factor *Motivation to Address Automatic Bias* ($d = 0.14$) showed the least amount of change compared to the other factors. It could be that although the bias education is effective in information transmission and persuasion, it may not inspire much motivation to do something about it. In addition, research shows that individuals are more likely to be goal motivated when they believe the effort will be successful (Weiner, 1985). The focus of the education was to teach people about automatic bias; it only briefly mentioned strategies that may help in reducing its effects. Therefore, participants may have left with a clear understanding that automatic biases exists and can influence behavior, but have been less sure of the ways to prevent those effects from happening.

In addition, the *Self-Perception of Objectivity* ($d = 0.25$) factor showed a weaker effect than the other factors. Interestingly, the education was effective in getting

participants to substantially agree that they had bias, but they did not show quite as strong of a disagreement in the parallel idea of being objective. This suggests that individuals find it easier to be convinced and agree that they have something undesirable (e.g., bias), but harder to persuade them to disagree that they have something desirable (e.g., objectivity).

Automatic Bias Education Differentially Impacts Social Groups

In addition to showing automatic bias education was effective; results also revealed that it impacts social groups differently. Females showed somewhat larger differences than males from pre to post-test on the factors *Automatic Biases Influence Me* and *Self-Perception of Objectivity*. Previous research shows that females are more motivated to respond without prejudice than males (Ratcliff, Lassiter, Markman, & Synder, 2006). Translating these findings to the present study, it could be that because women are more motivated to respond in nonprejudiced ways they are predisposed to believe that bias can influence them; as a result, receiving automatic bias education has a greater impact on their *beliefs about bias*.

In addition, Black participants showed some of the largest changes in their *beliefs about bias* from pre to post-test on the factors *I Have Bias*, *Automatic Bias Influences Me*, and *Self-Perception of Objectivity*. This initial finding suggests that bias education might impact Black participants differently than other groups. One possibility is that Black individuals have been taught about historical prejudices that emphasize how majority group members discriminate against Black people. In turn, Black participants may be less likely to believe that they could be biased since the cultural emphasis is on

how majority groups discriminate against them. Thus, receiving the bias education has greater impact for Black people, showing them that not only may they have bias toward individuals who are racially different than them, but they may also have biases toward those who are racially identical to them.

Receiving Automatic Bias Education Produced Durable Effects in the Short-Term

Notably, participants' changes in their *beliefs about bias* were durable in the short term. That is to say, the time from receiving the education to completing the questionnaire did not influence participants' responses (all p 's ≥ 0.12). Rather these effects were durable a few hours after receiving the education. A few hours is fine, but practical importance requires that such effects be durable for a longer time frame.

2.2 Study 2: Automatic Bias Education is Durable

The results from Study 1 suggest that bias education produces changes in *beliefs about bias* in the short term. Study 2 evaluated whether the changes in *beliefs about bias* persisted months after the intervention and whether it could predict change in social judgments in which biases could be expressed. In addition, Study 2 assessed whether the components of *beliefs about bias* relate to other social attitudes and beliefs. To assess whether the automatic bias education was durable, I re-examined participants' who completed Study 1 and compared their reported *beliefs about bias* to participants that did not receive the educational intervention.

Reviving the Effect of the Intervention with a Brief Reminder about Automatic Biases

Although the primary interest was in the durability of participants' change in their *beliefs about bias*, there are reasons to believe that durability may depend on making the

educational component accessible in memory first. So, one important test is to assess whether educational effects are durable with or without prompting. To test, I provided a very brief reminder about the education, but not its content, to a portion of the participants that had received the intervention. If effects are long-lasting with and without the prompt, this would suggest that reminding participants' about the education is unnecessary to illustrate the lasting effects. However, if participants' changes in *beliefs about bias* persist only when reminded about the bias education, it suggests that while the education can be quickly revived, it may not be readily accessible.

Does Automatic Bias Education also Change Self-Reported Attitudes, Automatic Attitudes, and Social Judgment?

It is possible that automatic bias education will not influence decisions or alter automatic processing, even if people's beliefs about bias have changed. According to Wilson and Brekke's mental contamination model (1994), simply being aware of one's unwanted thoughts is not enough to reduce the presence of bias. So even though Study 1 showed that the bias intervention was effective at making individuals aware of their bias, and providing them the motivation to address it (step 1 and 2 in the model) individuals still need to know the extent and direction of their bias and be able to eliminate its influence on their judgment or action. However, because the education had been given months prior, it is also possible that people used their motivation to seek out additional information regarding their biases to help them determine the extent and direction of their bias, giving them resources to alter its effects on judgment. Another possibility is that participants engaged in Monteith's self-regulation model (1993): because they became

more motivated to address bias, they might take more time when making a decision to ensure their biases do not influence them. For both of these cases, I would expect to see differences in social judgments and attitudes for individuals who received bias education compared to those who did not. To test whether education is a useful intervention for reducing bias, I examined a variety of dependent measures including self-reported racial attitudes, questions regarding bias, a race-IAT, and a judgment task in which racial biases could be expressed. The race-IAT and self-reported racial measures allowed me to assess whether education about bias had an effect on racial attitudes and automatic attitudes toward particular social groups or whether it only changed their beliefs about bias more generally. Given that the automatic bias education used examples highlighting racial bias, both the measures focused on attitudes toward Black and White individuals. The judgment task allowed me to examine whether participants who received the educational intervention and reminder relied less on stereotypical information. The task was a variation of the criminal versus politician task (Park & Banaji, 2000) that asks individuals to “remember” whether names connoting race (African American, European American) were criminals or not. Previous research has demonstrated that participants misidentify proportionally more African Americans as criminals than European Americans (Park & Banaji, 2000; Walsh, Banaji, & Greenwald, 1995).

Are Beliefs about Bias Correlated with Personality or Individual Difference Measures?

It could be that educating individuals about automatic bias is much like learning algebra or history where learning is fact-based, with individuals learning about the subject and readily applying the learnt material to novel situations. On the other hand, it

could be that there are individual differences in people's willingness to accept that automatic bias exists. That is to say, individuals may be differentially prepared, or motivated, to accept the possibility that biases may be automatic and that they have them whether they like it or not. To assess whether individual differences are related to people's *beliefs about bias*, I tested how the bias questionnaire related to the following measures: the Internal and External Motivation to Respond without Prejudice Scale (IMS, EMS; Plant & Devine, 1998), the Belief in a Just World Scale (BJW; Dalbert, Lipkus, Sallay, & Goch, 2001), and the Self-Monitoring Scale (Snyder, 1987). These three measures were chosen because I believed that they would correlate with people's willingness to accept that automatic bias exists and can influence behavior.

Belief in a Just World: General Just World Scale (BJW; Dalbert, Lipkus, Sallay, & Goch, 2001). The BJW scale measures how much individuals believe that people deserve what they get in life. Acknowledging automatic bias would imply that sometimes others are treated unfairly for reasons that they have no control over. Thus, I expected that participants who report high beliefs in a just world would show weaker *beliefs about bias* than individuals who report lower beliefs in a just world. Specifically, I expected that people who reported high beliefs in a just world to show less agreement on *I Have Bias*, *Automatic Bias Influences Me*, and *Motivation to Address Automatic Bias* than participants who reported lower beliefs in a just world. Likewise, I expected individuals who reported high beliefs in a just world to show more agreement on *Self-Perception of Objectivity* and *Knowing and Preventing Automatic Bias* than people who reported lower beliefs in a just world.

Internal and External Motivation to Respond without Prejudice Scale (IMS, EMS; Plant & Devine, 1998). The internal motivation to respond without prejudice scale (IMS) and external motivation to respond without prejudice scale (EMS) assess whether people's motivations for responding without prejudice is a result of their internal values or external social pressure. Given that one of the goals of Study 1 and 2 was to examine participants' willingness to acknowledge bias, I believed that participants' responses varied depending on their motivations to respond without prejudice. Importantly, people can report some combination of internal or external motivations. That is, a person could be both internally and externally motivated, only internally or externally motivated, or neither. As such, I made hypotheses about individuals' reports on the IMS and EMS scales separately.

There were two competing hypotheses for people with higher internal motivations to respond without prejudice. First, individuals who are internally motivated to identify those aspects that will ensure they act in-line with their values will recognize that they have biases that they are constantly working to overcome. In this case, participants who report high internal motivation to respond without prejudice would report stronger overall *beliefs about bias* than people who report low internal motivation to respond without prejudice. However, a competing hypothesis would be that because participants who score high on the IMS are internally motivated, they believe they accurately monitor themselves for bias and conclude that they are unbiased. In this case, they would show higher agreement on *Self Perception of Objectivity* and *Motivation to Address Automatic*

Bias, while reporting less agreement on the remaining factors than participants who score low on the IMS.

Participants who report high external motivation to respond without prejudice are concerned with appearing unprejudiced in front of others. As such, they should be more likely to report that they are objective and believe that addressing automatic bias is important. Thus, I expected participants with higher external motivations to respond without prejudice would report higher agreement on *Self-Perception of Objectivity* and the *Motivation to Address Automatic Bias* than participants who report lower external motivation to respond without prejudice. Similarly, participants with higher external motivations to respond without prejudice would be less likely to endorse that *Automatic Bias Influences Me*, *I Have Bias*, and *Knowing and Preventing Bias* than participants who report lower external motivation to respond without prejudice - as these factors focus mainly on individual responses (e.g., “Unwanted biases probably influence my decision making about other people”).

Self-Monitoring Scale (Snyder, 1987). The self-monitoring scale assesses how consistent an individual reports being across various situations. People who score high on this measure are more likely to adapt their behavior to coincide with the majority opinion, while individuals who score low are less likely to be concerned with others’ opinions and remain more behaviorally consistent across social situations. As such, I hypothesized that because high self-monitors modify their behavior to gain public acceptance, society’s emphasis on egalitarian values would make them more likely than low self-monitors report weaker *beliefs about bias*. In particular, I expected high self-monitors to show less

agreement on *I Have Bias* and *Automatic Bias Influences Me* than low self-monitors.

Likewise, I expected high-self-monitors to show more agreement on *Self-Perception of Objectivity*, *Motivation to Address Automatic Bias* and *Knowing and Preventing Automatic Bias* than low self-monitors.

Addressing Sampling and History Effects

This study utilized the psychology participant pool, selecting participants that did or did not receive the bias education from 3 to 6 months previously. Using the participant pool allowed me to minimize the extent to which individuals connected the study to the original intervention by (1) using a different setting, (2) recruiting “naturalistically” through the participant pool so that participants would not recognize any link to the original intervention, and (3) using different format of the questions intended to measure beliefs about bias.

There are several group or history effects that are important to consider. For example, those that are in the participant pool have presumably learned about particular subject areas of psychology, including additional information related to social biases. Thus, it was important to compare how much the effects are based on the educational intervention separate from the psychology class that they enrolled in.

Using the participant pool as the recruitment method also introduces a possible confounding factor – those that did not view the intervention are likely to be students entering their second year and beyond, making them potentially different than entering first year students in ways that are necessary to measure and rule out as alternatives. For example, first-year students in the participant pool may be more likely to pursue

psychology as a major, whereas those in later years may be more likely to be taking it as an elective. Moreover, participant pool students vary by which class they are currently enrolled in (e.g., social psychology students are likely to have more education about social and cognitive biases than are perception students). For example, one of the introductory classes exposes participants to the same themes as the bias education by the same lecturer (Brian Nosek) approximately two times a week.

I recruited an additional sample to directly address the potential history effects of selecting psychology and being in the participant pool. This sample was recruited around grounds rather than in the lab on a computer and included both participants in the treatment and control conditions. To ensure a large sample, I minimized the amount of time participants would need to complete the study. As such, this additional sample completed only the key dependent variable of interest for assessing durability effects of the intervention – the *beliefs about bias* questionnaire. Between the two sampling approaches, I can establish reasonable confidence whether differences observed between treatment and control groups can be attributed to the intervention.

2.2.1 Method

Participants

Laboratory Participants. 132 participants (61 male, 1 unreported, mean age = 19.02) from the University of Virginia psychology participant pool participated for partial course credit. Participants completed a 30 minute pre-selection questionnaire for the participant pool. One question asked “Did you see the first year lecture, "Mindbugs" that was presented in August 2010?” Participants that answered this question were eligible to

be in this study, and were divided into two groups: those that attended the education, and those who did not. Selection, in this case, is unobtrusive as these participants are “naturally” in the pool and participating in studies. However, there is an important selection effect – having chosen to take a psychology course. This might make the Study 2 subsample importantly different than the Study 1 full sample. Comparisons within Study 1 data for the subsample in Study 2 compared to the rest of the Study 1 sample will identify potential moderating influences of this selection effect. Also, the sample selecting to take psychology but that did not attend the bias education will be another useful comparison. Finally, for the participants who received the bias education, I have a longitudinal design. I can compare data from the pre- or post-test (Study 1) to their responses in this follow-up (Study 2).⁶

Of these participants, 12 (9.2%) reported being East Asian, 6 (4.6%) reported being South Asian, 18 (13.7%) reported being Black or African American, 85 (64.9%) reported being White, 2 (1.5%) reported being more than one race – Black/White, 5 (3.8%) reported being more than one race – other, 3 (2.3%) reported Other/Unknown, and 1 (0.01%) participant choose not to respond. 8 (6.0%) participants reported being of Hispanic or Latino ethnicity, 123 (93.2%) reported being not Hispanic or Latino origin, and 1 (0.8%) participants reported not knowing their ethnicity.

“Around Grounds” Participants. 207 participants (109 male, mean age = 19.78) from the University of Virginia volunteered to participate. Of these participants, 28 (13.5%) reported being East Asian, 11 (5.3%) reported being South Asian, 25 (12.1%)

---

⁶ There was an average of 120 days in-between Study 1 and 2 ($SD = 29.85$, range = 91-178 days). 90 participants were sampled during the Fall 2010 semester, 42 participants were sampled from the Spring 2011 participant pool.

reported being Black or African American, 112 (54.1%) reported being White, 3 (1.4%) reported being more than one race – Black/White, 18 (8.7%) reported being more than one race – other, 9 (4.3%) reported Other/Unknown, and 1 (0.01%) participant choose not to respond. 13 (6.3%) participants reported being of Hispanic or Latino ethnicity, 161 (77.8%) reported being not Hispanic or Latino origin, 10 (4.8%) participants reported not knowing their ethnicity, and 23 (11.1%) chose not to respond.

In addition, 64 (30.9%) participants indicated that they had attended the bias education, 121 (58.5%) indicated they had not received the bias education, and 22 (10.6%) indicated they were unsure whether they had received the education.

Design

Laboratory Participants. This study compared three groups in a between-subjects design: bias education participants who received the reminder, bias education participants who did not receive the reminder, and control participants who did not receive the educational intervention. For participants who received the bias education, there was also a longitudinal design, a mixed 2 (Reminder: Present versus Absent) X 2 (Time Assessment: Time 1 versus Time 2) design where one factor, Reminder, was between-subjects and the other factor, Time Assessment, was within-subjects.

To assess personal interest in the material and other selection effects, participants indicated (a) whether they were enrolled in the course as an elective or major requirement, (b) which course and instructor they had, (c) whether they were enjoying the course, and (d) whether they plan to enroll in more psychology courses in the future. Finally, participants were asked to report their university identification. Out of the

sample, 49% of the participants who received the bias education had provided me identification during Study 1; this allowed me to compare their data from Study 1 to this follow-up. For a complete list of participants' responses, refer to Table 7.

Table 7. Participants' Course Demographics

	Frequency
Reason for Enrolling in the Course	
Elective	11
Major Requirement	13
Sounded Interesting	105
Only One that Would Fit My Schedule	2
No Response	1
Course Enrolled	
Introductory Psychology	24 (Brian Nosek)
	16 (Chad Dodson)
	12 (Jon Bonvilian)
Cognition	26
Perception	1
Personality	22
Social Psychology	20
No Response	11
Are You Enjoying the Course?	
Yes	26
Planning on Being a Psychology Major?	
Definitely Yes	7
Probably Yes	12
Unsure	30
Probably No	37
Definitely No	46
Will You Enroll in More Psychology Courses?	
Definitely Yes	32
Probably Yes	33
Unsure	38
Probably No	24
Definitely No	5

“Around Grounds” Participants. Participants completed the bias questionnaire as well as basic demographics including sex, race, ethnicity, number of psychology courses taken, and year in college. In addition, participants indicated whether they received the education months earlier.

Materials

Bias Questionnaire. Participants recruited from the participant pool completed the same questionnaire as described in Study 1, with two differences. First, in lieu of a paper questionnaire, laboratory participants received a computerized version with each question presented one at a time and all possible response options vertically formatted, and simultaneously viewable. In addition, laboratory participants completed the other measures (e.g., interest questions) intermixed within the bias questions, making it less obvious that previous participants were completing the same questionnaire as completed months earlier. It is conceivable that these formatting modifications changed responding itself, but would be true for the laboratory comparison group as well.

In contrast, participants recruited “around grounds” completed the identical paper questionnaire as in Study 1. The results suggest that these formatting differences were immaterial.

Bias Education Reminder. All laboratory participants completed a 4-item questionnaire reporting which activities they had attended while enrolled. For the participants in the reminder condition, this questionnaire included a question asking whether they had attended the first year lecture, Mindbugs. This question was intended to remind participants of the automatic bias education lecture, and perhaps what they

learned would be more accessible. For control participants and the participants in the no-reminder condition, none of the items referred to the bias education lecture. For a complete list of questions, refer to Appendix B.

Participants recruited around grounds did not receive this manipulation making them most comparable to the no-reminder and control conditions.

Criminal Task (Park & Banaji, 2000). Laboratory participants were given White and Black sounding names and told that some may have been identified in public media as committing a crime. They were further instructed that although it would be unlikely that they would remember any of the details of the person or the crime, they may have a general sense of familiarity of having seen or heard the name previously. Participants were then asked to decide whether the individual was a criminal or not. The list consisted of 66 European American sounding names (e.g., John Olson) and 18 African American sounding names (e.g., Leroy Washington). These names were all generated for previous studies and were not associated with known criminals. The proportion of White and Black names was 79% and 21%, respectively. Participants were asked to read each name on the list and report if they judged the name to be that of a criminal. For the complete instructions and a complete list of names used, refer to Appendix C.

Implicit Association Test (IAT; Greenwald, et al. 1998) The IAT measures the strength with which concepts (e.g., Black and White people) are associated with attributes (e.g., good and bad). Stimulus items representing four categories are categorized one at a time and as quickly as possible using two keys of a computer keyboard. During the first critical response block *Black faces* and *good words*, for

example, are categorized with the same response key, while *White faces* and *bad words* are categorized with a second response key. In the second critical response block, the response pairings are reversed such that *Black faces* and *bad words* are categorized with one key, and *White faces* and *good words* with the other key. If the concepts sharing a response key are associated, it should be easier to categorize the stimulus items quickly compared to when the concepts sharing a response key are not associated (see Nosek, Greenwald, & Banaji, 2007 for a review). Laboratory participants completed one IAT measuring race associations with evaluation (good and bad). The specific IAT procedure followed recommendations of Nosek, Greenwald, and Banaji (2005), and scores were calculated with the *D* algorithm (Greenwald, Nosek, & Banaji, 2003).

Self-Reported Racial Attitudes. Laboratory participants reported their self-reported feelings toward Black people and White people using a 7-point scale (1=strongly prefer White people to Black people, 7=strongly prefer Black people to White people). In addition, participants reported their explicit warmth rating toward Black and White individuals using a 10-point Likert scale (1=cold, 10=warm).

Belief in a Just World Scale: General Just World Scale (Dalbert, Lipkus, Sallay, & Goch, 2001). Laboratory participants responded to questions related to their beliefs that the world is a just place. Participants reported their beliefs on a (1=strongly disagree, 6=strongly agree). For a complete list, refer to Appendix D.

Internal and External Motivation to Respond without Prejudice Scale (Plant & Devine, 1998). Laboratory participants responded to questions concerning their motivations to respond without prejudice using a 6-point scale (1=strongly disagree,

6=strongly agree). In lieu of targeting a particular social group, I generalized the questionnaire to reflect general prejudice. For example, instead of being asked, “I attempt to act in nonprejudiced ways toward *Black people* because it is personally important to me,” participants would be asked “I attempt to act in nonprejudiced ways toward *other people* because it is personally important to me.” For a complete list, refer to Appendix E.

Self-Monitoring (Snyder, 1987). The self-monitoring scale distinguishes between people who monitor their behavior to fit different situations (i.e., high self-monitors), and people whose behavior remains fairly consistent across various situations (i.e., low self-monitors). Laboratory participants responded to questions relating to how they act in the presence of other on a 6-point scale (1=strongly disagree, 6=strongly agree). For a complete list, refer to Appendix F.

Procedure

Laboratory Participants. Participants were recruited via an on-line study participation system. Participants did not know that their responses on pre-selection were relevant for their eligibility in the study, nor did they have any information about study purposes until they were actually in the study itself. After arriving at the laboratory and signing informed consent, all participants reported which university activities they had attended and completed the bias questionnaire with demographics questions randomly intermixed. Afterwards, participants were told that they would be taking part in another, unrelated study. To add credibility, participants were escorted to another room down the hall and were introduced to a new experimenter who provided the instruction and materials for the remaining measures. There, participants completed the judgment task,

the individual difference measures, and a randomized order of implicit and explicit race measures.

"Around Grounds" Participants. Participants were recruited around grounds. A research assistant (RA) approached them, asking if they were currently enrolled as a student. If they were, the RA asked whether they would be willing to complete a five minute psychology survey. After receiving informed consent, the participants completed only the bias questionnaire, basic demographic questions, and reported if they had seen the bias lecture at the beginning of the academic year.

2.2.2 Results and Discussion

Results

For all dependent measures, an ANOVA was used first to test whether there were sampling effects that might have elicited differences in beliefs about bias. The following variables were included in the model: which class they were enrolled in, which psychology instructor they had, whether they were enjoying the course, and their reported reason (i.e., major versus elective) for enrolling in the course. Results revealed no significant differences for the any of the measures on their overall BaB score (Reason for Enrolling in the Course, $p = 0.60$, Course Enrolled In, $p = 0.16$, Enjoying the Course, $p = 0.22$, Planning on Becoming a Psychology Major, $p = 0.21$, Planning on Enrolling in More Psychology Courses, $p = 0.24$). Results also revealed no significant differences on any of the subscales. Therefore, these measures were not included in any of the analyses described below.

Does Being Reminded About the Education Matter?

Laboratory participants that had received the intervention were randomly assigned to get a reminder of having seen the education or not. If this reminder matters, then it will be important to consider the effects of these two subgroups separately. If it did not matter, then I can increase power of subsequent tests by collapsing these two conditions for the remainder of the analyses.

A *t*-test revealed that participants who received the reminder did not show significantly different overall BaB scores compared to those who did not, $t(92) = 0.95$, $p = 0.34$, $d = 0.20$. Likewise, participants who received a reminder about receiving the education did not show greater belief in *I Have Bias*, $t(92) = 0.60$, $p = 0.55$, $d = 0.13$, *Automatic Bias Influences Me*, $t(92) = 0.10$, $p = 0.92$, $d = 0.02$, *Self-Perception of Objectivity*, $t(92) = 1.18$, $p = 0.25$, $d = 0.04$, *Motivation to Address Automatic Bias*, $t(92) = 0.03$, $p = 0.98$, $d = 0.01$, or *Knowing and Preventing Automatic Bias*, $t(92) = 1.58$, $p = 0.12$, $d = 0.33$, than participants who did not receive the reminder. Because providing a reminder did not qualify any of the results for participants' beliefs, I combined the two conditions for the remaining analyses.

Longitudinal Effects - Examining Change in Automatic Bias Education Participants Beliefs over Time

Of the 94 participants who completed both Study 1 and 2, were 61 sampled from the Fall semester and 33 were sampled from the Spring semester. Of these 94 participants, 45 provided sufficient identification at Time 1 and Time 2 to link their scores. The longitudinal analysis described below uses this subsample to test for durability longitudinally.

To test for longitudinal effects, I evaluated whether the differences that (may) have been observed in Study 1 persisted months later using a 2 (Time Measured: Pre versus Post at Time 1) X 2 (Test Assessment: Time 1 and Time 2) mixed-design ANOVA to see if there are changes in participants' beliefs that unwanted biases can influence their decisions. One factor – pre/post at Time 1 was a between-subjects factor while the factor Time 1 and Time 2 is a within subjects factor. For a graphical representation of participants' *beliefs about bias* at Time 1 and Time 2, refer to Figure 1 below.

Figure 1. Bias educated participants' scores at Time 1 and Time 2.

If the bias education produced durability, then participants who reported their beliefs about bias prior to the presentation should show stronger beliefs about bias at Time 2, and those that reported their beliefs about bias after the presentation might show little or no decline in beliefs about bias at Time 2. There was not a significant interaction of Time Measured X Test Assessment on participants' overall *beliefs about bias* (BaB) score, $F(1, 43) = 2.09, p = 0.16, d = 0.44$, though the effect size was sizable. However, results revealed a main effect of time $F(1, 43) = 24.49, p < 0.0001, d = 0.75$, such that, on

average, participants' *beliefs about bias* increased from Time 1 ($M = 3.60$, $SD = 0.53$) to Time 2 ($M = 3.96$, $SD = 0.38$).

Testing each group individually, a t -test revealed that participants who completed the questionnaire prior to the education at Time 1 ($M = 3.47$, $SD = 0.54$) showed a significant increase in their BaB score at Time 2 ($M = 3.91$, $SD = 0.38$), $t(25) = 4.32$, $p = 0.0002$, $d = 0.86$. Results also showed participants who completed the questionnaire after receiving the education at Time 1 showed a significant *increase* in their BaB score from Time 1 ($M = 3.79$, $SD = 0.46$) to Time 2 ($M = 4.02$, $SD = 0.38$), $t(18) = 3.01$, $p = 0.01$, $d = 0.71$. Despite having low power, the results suggest that the intervention is durable.

I Have Bias. Results showed that participants who completed the bias questionnaire prior to the presentation did not show greater change from Time 1 to Time 2 compared to those completing the questionnaire afterwards, $F(1, 43) = 0.60$, $p = 0.44$, $d = 0.24$. However, results did reveal a main effect of time $F(1, 43) = 23.78$, $p < 0.0001$, $d = 0.74$ such that participants were more likely to endorse *I Have Bias* at Time 2 ($M = 4.69$, $SD = 0.75$) than at Time 1 ($M = 4.13$, $SD = 1.02$).

Upon closer examination, a t -test showed that participants who completed the questionnaire prior to the education at Time 1 ($M = 4.01$, $SD = 1.12$) showed a significant increase in their score at Time 2 ($M = 4.64$, $SD = 0.80$), $t(25) = 3.78$, $p = 0.001$, $d = 0.76$, suggesting stronger beliefs that they could have bias. Likewise, participants who completed the questionnaire after receiving the education at Time 1 ($M = 4.29$, $SD = 0.86$) showed a significant increase in their score at Time 2 ($M = 4.75$, $SD = 0.70$), $t(19) = 3.62$, $p = 0.002$, $d = 0.83$.

Automatic Bias Influences Me. Results showed that participants who completed the bias questionnaire prior to the presentation did not show greater change from Time 1 to Time 2 compared to those completing the questionnaire afterwards, $F(1, 43) = 0.91, p = 0.34, d = 0.29$. There was, however, a significant main effect of time showing that participants' were more likely to believe that *Automatic Bias Influences Me* at Time 2 ($M = 4.29, SD = 0.75$) than they were at Time 1 ($M = 3.83, SD = 0.97$), $F(1, 43) = 8.76, p = 0.01, d = 0.45$.

Following-up on this result, a t -test revealed that participants who completed the questionnaire prior to the education at Time 1 ($M = 3.58, SD = 1.11$) showed a significant increase in their beliefs that *Automatic Bias Influences Me* at Time 2 ($M = 4.16, SD = 0.82$), $t(25) = 2.58, p = 0.02, d = 0.52$. Participants who completed the questionnaire after receiving the education at Time 1 ($M = 4.18, SD = 0.60$) did not show significant change in their score at Time 2 ($M = 4.47, SD = 0.62$), $t(18) = 1.88, p = 0.08, d = 0.44$, though the effect size was relatively large.

Self-Perception of Objectivity. Results revealed that participants who completed the bias questionnaire prior to the presentation did not have greater change from Time 1 to Time 2 compared to those completing the questionnaire afterwards, $F(1, 43) = 2.57, p = 0.12, d = 0.49$. In addition, results revealed no significant main effect of time $F(1, 43) = 2.05, p = 0.16, d = 0.22$, suggesting that participants' beliefs in *Self-Perception of Objectivity* did not significantly change from Time 1 ($M = 5.32, SD = 0.71$) to Time 2 ($M = 5.07, SD = 0.85$).

However, individual level t -tests revealed that participants who completed the questionnaire prior to the education showed significantly lower beliefs in their objectivity at Time 2 ($M = 4.94$, $SD = 0.97$) than they did at Time 1 ($M = 5.40$, $SD = 0.75$), $t(25) = 2.31$, $p = 0.03$, $d = 0.46$. In contrast, results showed that participants who completed the questionnaire after receiving the education showed no significant change in their *Self-Perception of Objectivity* at Time 2 ($M = 5.21$, $SD = 0.65$) than at Time 1 ($M = 5.23$, $SD = 0.63$), $t(18) = 0.12$, $p = 0.91$, $d = 0.03$. This pattern of effects suggests durability in the expected form – change only observed for the group measured prior to the original intervention.

Motivation to Address Automatic Bias. Results showed that participants who completed the bias questionnaire prior to the presentation did not have greater change from Time 1 to Time 2 compared to those completing the questionnaire afterwards, $F(1, 43) = 0.02$, $p = 0.88$, $d = 0.04$. Results also revealed no significant main effect of time $F(1, 43) = 0.39$, $p = 0.54$, $d = 0.10$, suggesting that participants' *Motivation to Address Automatic Bias* did not significantly change from Time 1 ($M = 3.97$, $SD = 0.63$) to Time 2 ($M = 3.88$, $SD = 0.66$).

A t -test showed that participants who completed the questionnaire prior to the education at Time 1 ($M = 4.00$, $SD = 0.64$) showed no significant change in their *Motivation to Address Automatic Bias* at Time 2 ($M = 3.90$, $SD = 0.46$), $t(25) = -0.68$, $p = 0.50$, $d = 0.14$. Results also revealed that participants who completed the questionnaire after receiving the education at Time 1 ($M = 3.92$, $SD = 0.62$) did not show significant

change in their motivation at Time 2 ($M = 3.85$, $SD = 0.88$), $t(18) = 0.27$, $p = 0.79$, $d = 0.06$.

Knowing and Preventing Bias. Results showed that participants who completed the bias questionnaire prior to the presentation did not show greater change from Time 1 to Time 2 compared to those completing the questionnaire afterwards, $F(1, 43) = 0.39$, $p = 0.54$, $d = 0.19$. However, results showed a main effect of time such that participants reported less agreement that at *Knowing and Preventing Automatic Bias* (the predicted direction of change) at Time 2 ($M = 3.39$, $SD = 0.90$) than they had during Time 1 ($M = 4.11$, $SD = 0.95$), $F(1, 43) = 18.91$, $p < 0.0001$, $d = 0.66$.

Follow-up t -tests revealed that participants who completed the questionnaire prior to the education at Time 1 ($M = 4.35$, $SD = 0.85$) showed a significant decrease in their belief in *Knowing and Preventing Automatic Bias* at Time 2 ($M = 3.54$, $SD = 0.76$), $t(25) = 3.76$, $p = 0.001$, $d = 0.75$. A t -test also revealed that participants who completed the questionnaire after receiving the education at Time 1 ($M = 3.79$, $SD = 1.00$) showed significant decrease in their belief at Time 2 ($M = 3.18$, $SD = 1.06$), $t(18) = 2.52$, $p = 0.02$, $d = 0.59$.

Overall, the results suggest that, despite low power, a subset of participants' *beliefs about bias* were durable over time. In particular, participants reported stronger agreement at Time 2 that they had bias and that automatic bias could influence them than they did at Time 1. Participants also reported less agreement that they would know and be able to prevent automatic bias at Time 2 compared to Time 1.

However, a longitudinal design like this does not guarantee that the intervention is the cause of the lasting change. Between Time 1 and Time 2 participants may have had other experiences that impacted their *beliefs about bias*. To assess possible history effects, I needed to compare this group's effects with control participants that did not receive the intervention, but were in the participant pool (effects of taking psychology courses), and with intervention and control participants that were not in the participant pool (individual differences in selecting psychology and history effects more generally). For these comparisons, Time 1 data is irrelevant, so I only examined data from Time 2.

Between-Groups Effects – Comparing People's Beliefs about Bias that Did or Did Not Receive Automatic Bias Education

To examine whether participants' *beliefs about bias* changed as a function of the education, I used a 2 (Sample: Laboratory versus "Around Grounds") X (Treatment: Intervention versus Control) ANOVA to test whether participants who received the education reported significantly different *beliefs about bias* than participants who had not received the education.⁷ For a complete list of means, refer to Table 8.

If the bias education is durable, then educated participants should show stronger beliefs than those who were not. But, if the apparent durability of beliefs about bias in the longitudinal analysis above were due to history effects, then these groups may not differ. I could also assess potential sampling effects to see whether the laboratory participants

---

⁷ It is worth noting that Study 2 control participants showed stronger overall beliefs about bias ($M = 4.00$, $SD = 0.45$) than participants who reported their beliefs prior to receiving the education at Time 1 ($M = 3.60$, $SD = 0.51$), $t(1187) = 9.34$, $p < 0.0001$, $d = 0.54$. This is additional evidence for a history effect, perhaps due to formative changes during the college experience.

were different than those recruited “around grounds” – perhaps because of selecting to take a psychology course.

Table 8. Means for the Between-Groups Comparisons

	Laboratory				“Around Grounds”			
	Treatment		Control		Treatment		Control	
	Mean	St Dev	Mean	St Dev	Mean	St Dev	Mean	St Dev
Overall BaB Score	4.01	0.42	3.93	0.41	4.16	0.51	4.02	0.46
I Have Bias	4.71	0.78	4.60	0.79	4.94	0.78	4.64	0.79
Automatic Bias Influences Me	4.36	0.76	4.14	0.70	4.51	0.83	4.25	0.73
Self-Perception of Objectivity	4.91	0.83	4.80	0.82	4.98	0.75	4.81	0.92
Motivation to Address Automatic Bias	4.77	1.01	4.66	0.72	5.05	1.02	4.86	0.96
Knowing and Preventing Automatic Bias	3.41	0.99	3.47	0.99	3.36	1.05	3.46	1.05

Results revealed that there was not a significant interaction of Sample X

Treatment on participants’ overall *beliefs about bias* (BaB) score, $F(1, 313) = 0.30, p = 0.58, d = 0.06$. However, results showed a main effect of treatment such that participants who received the education reported higher overall BaB scores ($M = 4.07, SD = 0.47$) than participants who had not received it ($M = 4.00, SD = 0.45$), $F(1, 313) = 3.86, p = 0.05, d = 0.22$. Also, results revealed a main effect of sample suggesting that participants sample “around grounds” had higher BaB scores ($M = 4.07, SD = 0.48$) than laboratory participants ($M = 3.98, SD = 0.42$), $F(1, 313) = 5.04, p = 0.03, d = 0.25$.

I Have Bias. Results revealed no significant interaction of Sample X Treatment on participants’ reported belief on *I Have Bias*, $F(1, 313) = 0.89, p = 0.35, d = 0.11$. There

was a main effect of treatment such that participants who received the education reported more agreement that they had bias ($M = 4.80$, $SD = 0.78$) than participants who had not received it ($M = 4.63$, $SD = 0.79$), $F(1, 313) = 4.39$, $p = 0.04$, $d = 0.24$. In addition, results revealed no main effect of sample such that participants sampled “around grounds” did not report more agreement that they had bias ($M = 4.75$, $SD = 0.80$) than laboratory participants ($M = 4.68$, $SD = 0.78$), $F(1, 313) = 1.99$, $p = 0.16$, $d = 0.16$.

Automatic Bias Influences Me. Results revealed no significant interaction of Sample X Treatment on participants’ reported belief on *Automatic Bias Influences Me*, $F(1, 313) = 0.06$, $p = 0.81$, $d = 0.03$. There was a main effect of treatment such that participants who received the education were more likely to agree that automatic bias could influence them ($M = 4.42$, $SD = 0.79$) than participants who had not received it ($M = 4.22$, $SD = 0.73$), $F(1, 313) = 6.55$, $p = 0.01$, $d = 0.29$. Results also revealed no main effect of sample; participants sampled “around grounds” did not report more agreement that automatic bias could influence them ($M = 4.34$, $SD = 0.78$) than laboratory participants ($M = 4.30$, $SD = 0.75$), $F(1, 313) = 1.93$, $p = 0.17$, $d = 0.16$.

Self-Perception of Objectivity. Results revealed no significant interaction of Sample X Treatment on participants’ reported belief on *Self-Perception of Objectivity*, $F(1, 313) = 0.08$, $p = 0.78$, $d = 0.03$. In addition, results revealed no main effect of treatment; participants who received the education reported similar beliefs in their objectivity ($M = 4.94$, $SD = 0.80$) as participants who had not received it ($M = 4.81$, $SD = 0.90$), $F(1, 313) = 1.80$, $p = 0.18$, $d = 0.15$. Finally, results revealed no main effect of sample; participants sampled “around grounds” were not more or less likely to believe

they were objective ($M = 4.87$, $SD = 0.87$) than laboratory participants ($M = 4.88$, $SD = 0.83$), $F(1, 313) = 0.15$, $p = 0.70$, $d = 0.04$.

Motivation to Address Automatic Bias. Results revealed no significant interaction of Sample X Treatment on participants' reported belief on *Motivation to Address Automatic Bias*, $F(1, 313) = 0.13$, $p = 0.72$, $d = 0.04$. Results also showed no main effect of treatment; participants who received the education showed similar beliefs in their objectivity ($M = 4.89$, $SD = 1.02$) as participants who had not received it ($M = 4.81$, $SD = 0.91$), $F(1, 313) = 1.74$, $p = 0.19$, $d = 0.15$. Interestingly, results revealed a main effect of sample; participants sampled "around grounds" reported more agreement on *Motivation to Address Automatic Bias* ($M = 4.92$, $SD = 0.98$) than laboratory participants ($M = 4.74$, $SD = 0.93$), $F(1, 313) = 4.11$, $p = 0.04$, $d = 0.23$.

Knowing and Preventing Automatic Bias. Results revealed no significant interaction of Sample X Treatment on participants' reported belief on *Knowing and Preventing Automatic Bias*, $F(1, 313) = 0.03$, $p = 0.86$, $d = 0.02$. Results also revealed no main effect of treatment; participants who received the education reported similar beliefs in their objectivity ($M = 3.39$, $SD = 1.01$) as participants who had not received it ($M = 3.47$, $SD = 1.03$), $F(1, 313) = 0.47$, $p = 0.49$, $d = 0.08$. In addition, results did not reveal a main effect of sample suggesting that participants sampled "around grounds" were not more or less likely to agree that they would know or be able to prevent automatic bias ($M = 3.43$, $SD = 1.05$) than laboratory participants ($M = 3.43$, $SD = 0.99$), $F(1, 313) = 0.06$, $p = 0.49$, $d = 0.03$.

Did Bias Education Influence Participants Automatic and Self-Reported Racial Attitudes?

Because “around grounds” participants did not complete any of the additional measures, they are not included in analyses for the remaining dependent measures. Instead, the results reflect *t*-test comparisons of laboratory treatment and control groups.

Automatic Racial Attitudes. It is possible that the previous education made participants more aware of their automatic attitudes and made them more likely to consciously try to override them. Overall, participants implicitly preferred White people to Black people ($M = 0.40$, $SD = 0.41$), $t(129) = 11.21$, $p < 0.01$, $d = 0.99$. If the previous education made participants more aware of their automatic attitudes and made them more likely to develop strategies to change or override them, then participants who received the bias education might have less implicit pro-White bias compared to those who did not. This did not occur. A *t*-test revealed no difference between individuals who received the bias education and control participants, $t(128) = 0.05$, $p = 0.96$, $d = 0.01$.

Self-Reported Racial Attitudes. Overall, participants explicitly preferred White people to Black people ($M = 0.59$, $SD = 1.03$), $t(126) = 6.44$, $p < 0.001$, $d = 0.57$. Paired *t*-tests revealed that participants reported significantly higher feelings of warmth toward White people ($M = 7.65$, $SD = 1.72$) than Black people ($M = 5.52$, $SD = 2.27$), $t(127) = 8.35$, $p < 0.001$, $d = 0.74$. A *t*-test revealed no significant differences between intervention and control conditions for participants’ explicit preference, $t(125) = 0.97$, $p = 0.33$, $d = 0.09$, feelings of warmth toward White people, $t(126) = 0.47$, $p = 0.64$, $d = 0.04$, or feelings of warmth toward Black people, $t(126) = 0.76$, $p = 0.44$, $d = 0.07$. For a

complete list of means of automatic and self-reported racial attitudes by conditions, refer to Table 9.

Table 9. Participants' Automatic and Self-Reported Racial Attitudes by Condition

	Mean	St Dev
Automatic Racial Attitudes*		
Bias Education Participants	0.40	0.44
Control Participants	0.40	0.33
Explicit Preference**		
Bias Education Participants	0.53	1.07
Control Participants	0.73	0.93
Feelings of Warmth toward Whites***		
Bias Education Participants	7.52	1.70
Control Participants	7.68	1.78
Feelings of Warmth toward Blacks***		
Bias Education Participants	5.42	2.33
Control Participants	5.76	2.14

*higher scores indicate greater automatic preference for White people compared to Black people

**this scale is centered at zero with positive values indicating more explicit preference for White people compared to Black people

***higher scores indicate warmer feelings toward Black and White individuals

Did Automatic Bias Education Influence Biases in Social Judgments?

It is possible that the previous education reduced bias in participants' social judgments such that educated participants would be less likely to judge Black sounding names as belonging to a criminal than non-bias educated participants. Overall, laboratory participants judged 33.3% of Black sounding names (e.g., Leroy Washington) as belonging to a criminal compared to 29.0% of White sounding names (e.g., John Olson). For a complete list of the percentage of participants, by condition, who judged Black and White sounding names as belonging to a criminal, refer to Table 10.

A 2 (Target Race: Black versus White) X 2 (Condition: Bias Education versus Control) ANOVA was conducted on the proportions of names identified as criminals. Across conditions, Black sounding names were more likely than White sounding names to be misjudged as

criminals, $F(1, 127) = 5.29, p = 0.02, d = 0.20$. Results revealed no significant interaction of Target Race X Condition for participants' criminal judgments of Black and White sounding names, $F(2, 127) = 0.48, p = 0.62, d = 0.12$ indicating that the tendency to misjudge Black sounding names as criminal did not decrease as a function of receiving bias education.

Individual Difference Measures - Examining Whether the Components of Beliefs about Bias Relate to Other Beliefs

Table 11 shows that in contrast to my hypotheses, participants' *beliefs about bias* were not significantly correlated with internal or external motivations to respond without prejudice, beliefs in a just world, or self-monitoring.

For a completed list of the means (overall and by condition) of the individual difference measures, refer to Table 12.

Table 10. Percentage of "Yes" Judgments

	Mean	St Dev
Bias Education Participants		
Black Names	33.2	25.7
White Names	30.1	18.3
Control Participants		
Black Names	33.6	27.3
White Names	26.4	18.2

Table 11. Correlation between Individual Difference Scores and Beliefs about Bias

	Overall BaB Score	I Have Bias	Automatic Bias Influences Me	Self-Perception of Objectivity	Motivation to Address Automatic Bias	Knowing and Preventing Automatic Bias
Belief in a Just World	-0.07	-0.01	-0.08	-0.03	-0.05	0.08
Internal Motivation to respond Without Prejudice	-0.04	-0.01	-0.09	0.02	0.10	0.18*
External Motivation to Respond Without Prejudice	-0.05	-0.02	-0.11	-0.04	0.02	0.00
Self-Monitoring	-0.03	0.02	0.00	0.08	0.10	0.09

* $p < 0.05$

Table 12. Individual Difference Measure Means

	Mean	St Dev
Belief in a Just World		
Overall	3.44	0.79
Bias Education Participants	3.65	0.61
Control Participants	3.50	0.86
IMS		
Overall	4.70	0.94
Bias Education Participants	4.81	0.91
Control Participants	4.42	0.96
EMS		
Overall	3.91	1.00
Bias Education Participants	3.80	1.03
Control Participants	4.18	0.88
Self-Monitoring		
Overall	3.29	0.60
Bias Education Participants	3.42	0.77
Control Participants	3.50	0.86

Discussion

Study 2 showed that participants' changes regarding some of their *beliefs about bias* persisted months after the intervention. Also, the results showed that the components of *beliefs about bias* were not related to other social attitudes and beliefs. Finally, participants who received the education showed similar bias in their social attitudes and judgments as participants who did not receive the education.

Automatic Bias Education – Regardless of Receiving a Reminder – Showed Durability on Individuals' Beliefs about Bias

Examining participants' *beliefs in bias* longitudinally, results provide evidence that the bias education was durable. Interestingly, some of participants' *beliefs about bias* increased over time. BaB scores were higher on average at Time 2 compared to both groups – tested before or after the intervention – at Time 1. Participants agreed more strongly at Time 2 that they had bias and that automatic bias could influence them than they did at Time 1. In addition, participants reported less agreement that they would know and be able to prevent automatic bias at Time 2 compared to Time 1. This analysis on its own, however, does not rule out history effects that could have influenced participants' beliefs about bias during the delay, regardless of whether they received the bias education.

The between-subject comparisons provide evidence that there are some history effects as well as a persisting influence of intervention. This suggests that some of participants' *beliefs about bias* changed over time regardless of receiving the intervention or not. Control participants' reported significantly stronger overall *beliefs about bias* (*M*

= 4.00, $SD = 0.45$) than participants who completed the questionnaire prior to receiving the education at Time 1 ($M = 3.46$, $SD = 0.54$). However, those control participants still more weakly endorsed beliefs about bias than the educated participants at Time 2. This combination illustrates an effect of history and of the intervention. As students begin their academic careers they may be getting experiences that influence their beliefs about bias. For example, participants may interact with a more diverse population than they had previously, or may learn about diversity and bias through coursework. Indeed, research shows that college students become more tolerant and open to change during their collegiate experience (see Pascarella et al. 1996 for a review). Notably, the history effect appears to be stronger than the durable influence of the intervention. This may be unsurprising as the intervention was a single 1-hour event, whereas the collegiate experience is a daily acculturation. Nevertheless, participants who received the bias education reported higher beliefs on some of the factors (i.e., *I Have Bias* and *Automatic Bias Influences Me*) than control participants.

In addition to the education's durability, the results also showed that for some of the *beliefs about bias*, "around grounds" reported significantly stronger agreement on *Motivation to Address Bias* than laboratory participants. Interestingly, this was the only sub-component of participants' *beliefs about bias* that showed a significant difference between the two samples. One possible explanation for these results is procedural: "around grounds" participants reported their beliefs in front of the experimenter whereas laboratory participants reported their beliefs on a computer. Thus "around grounds" participants may have experienced greater self-presentational concerns, which I presume

would be more likely to influence reported “motivations” than descriptive beliefs about how bias operates.

People’s Willingness to Believe in Bias did not Correlate with Conceptually Related Individual Difference Measures

Results revealed that the individual difference measures (Belief in a Just World, Internal and External Motivation to Respond without Prejudice, Self-Monitoring) were not significantly correlated with people’s beliefs that automatic biases exist and influence behavior. One possible explanation is that participants who are similar on one personality dimension respond differently on their *beliefs about bias*. For example, high self-monitors might show strong *and* weak *beliefs about bias*. On one hand, because high self-monitors care about what others think, society’s emphasis on egalitarian values might lead them to report strong *beliefs about bias*. However, other high self-monitors might believe that the admission of having bias would make others look down on them; as a result, they report weaker *beliefs about bias*. Using the factor *I Have Bias* as an illustration, a similar case could be made for participants who score high on internal motivation to respond without prejudice. Individuals who are internally motivated might believe that they have biases that they are constantly working to overcome. In this case, participants might report strong agreement on *I Have Bias*. However, other high internally motivated people might believe that because they are constantly monitoring themselves to ensure objectivity, they do not harbor bias. Thus, they would disagree with the factor *I Have Bias*.

Another possibility is that participants' themselves do not perceive a relationship between their personal beliefs about the world and bias; as a result, their *beliefs about bias* may reflect variation in their knowledge (e.g., intuition, societal norms, experience) that is not a motivated belief about how the world works. Likewise, their responses may be descriptive beliefs about how bias works, rather than the motivations or values about what one should do about it. Thus, participants might respond in ways such that agreement or disagreement regarding their *beliefs about bias* does not correspond with any particular motivation, personality, or ideology about how the world works.

Alternatively, it could be that participants' *beliefs about bias* are indicative of personality traits that are distinct from these three individual difference measures. If, after further examination, participants' *beliefs about bias* remain uncorrelated to other existing individual difference measures, then future research may reveal that people's *beliefs about bias* are predictive and connected to some, not yet identified, motivation or belief systems.

Automatic Bias Education Did Not Influence Participants' Social Attitudes or Judgments

Two important questions to examine were whether automatic bias education could influence judgments in which bias could be expressed, and whether it could promote racial attitude change. Results revealed that it did not. Participants who received automatic bias education showed similar automatic and self-reported racial attitudes compared to laboratory control participants. Likewise, participants who received the automatic bias education showed similar levels of stereotyping in social judgments as laboratory participants who did not receive this education.

These results may not be very diagnostic as to whether automatic bias can influence social judgments. For one, although there was some evidence for durability, it may have been too weak to affect individuals' social judgments measured months after the intervention. Testing individuals more closely in time after receiving the education may provide a better understanding of whether the intervention can influence such judgments, at least temporarily. Nevertheless, it is possible that receiving bias education will not be enough to change participants' judgments in which bias could be expressed either in the short or long-term; Wilson and Brekke's model (1994) states that in order for participants' to effectively alter their response, they must know the extent their bias, and know how to undo its influence. These were not addressed in the intervention. Thus, it is possible that the intervention is only effective at addressing the first two stages of the model – awareness of having bias and motivation to address it – without affecting the ability to actually change their behavior.

2.3 Study 3: Automatic Bias Education Compared to Other Interventions

Study 1 and 2 examined whether bias education could influence individuals' *beliefs about bias*, and Study 2 also included measures of social attitudes and judgments. These studies took advantage of a naturally occurring event providing ecological validity, conducted longitudinal tests for examining durability, and had some experimental control in the initial session assisting with internal validity. However, the studies did not make any comparisons with other ways of learning about bias.

In Study 3, I sought to maximize internal validity by bringing full experimental control into the laboratory, and to test whether the *experiential* component of the bias

education has impact on beliefs and social judgment beyond other successful interventions that educate individuals about bias. Thus, I experimentally compared a distilled version of the bias education with two other effective bias interventions: (1) Pronin and Kugler's (2007) bias blindspot exercise, and (2) simple exposure to the IAT without accompanying experiential education (Menatti et al., 2011; Monteith, et al, 2001; Monteith et al., 2002). This sets-up a tough test for the automatic bias education as these interventions incorporate components of the same ideas, and existing evidence supports their effectiveness.

Another challenge is that to bring the bias education into the lab I needed to simplify it dramatically from the in-person, large lecture format from Study 1. It is possible that I would make changes that would eliminate some important components of the intervention – i.e., the full hour of education, more direct experience with the demonstrations, and the social influence of experiencing the education with other audience members. Even so, I hypothesized that distilling the intervention to the key components and delivering it with a video in private would produce meaningful change in *beliefs about bias* and possibly on social judgments and attitudes.

To test, participants were randomly assigned to a control condition or one of three interventions: a shortened video clip of the original bias education plus completing the IAT, the bias blindspot education (Pronin & Kugler, 2007), or just the IAT with no additional education. After receiving one of the interventions, participants completed various questionnaires regarding their *beliefs about bias*, a judgment task, and a measure of behavioral intentions/policy support about bias.

Comparison Intervention 1: Bias Blindspot Education

Pronin and Kugler's (2007) intervention presents a scientific article summarizing evidence that situational factors – some that individuals may be unaware of – can influence judgments and behavior. For example, participants read about the bystander effect (Darley & Latané, 1968), a phenomenon in which the greater number of people present reduces the likelihood that any one person will help. To assess whether the intervention was effective, Pronin and Kugler asked participants to indicate how much, compared to their peers, they would be susceptible to the specific biases they had previously learned about. Participants who read the article were more likely to admit that they were susceptible to these situational biases than control participants.

The present study also included Pronin and Kugler's (2007) questionnaire in addition to the *Beliefs about Bias* measure employed in the previous two studies. This helped ascertain whether the bias interventions produced changes specific to their respective dependent measures, or if the interventions produced general changes in beliefs about bias and displayed effects on either dependent measure.

Comparison Intervention 2: Experience with the IAT

The bias intervention includes the IAT as part of its experiential education. Some research demonstrates that the IAT on its own is sufficient to change subsequent attitudes. For example, Menatti, Smyth, Teachman, and Nosek (2011) showed that participants who completed an IAT to assess their automatic bias toward people with mental illness prior to self-report showed reduced explicit stigma compared to participants who completed the IAT after self-report. This suggests that the completing

the IAT made participants aware of potential bias, and motivated correction for it (see also Monteith et al., 2001).

Will Bias Education Produce Stronger Effects than the comparison interventions?

Because these interventions are comparable to the bias educating by teaching individuals about automatic biases either through knowledge or experience, it is possible that the bias education will elicit similar effects on individuals' *beliefs about bias*. Notably, the pragmatics of testing these interventions in a well-controlled context required that I reduce both the time and some of experiential features of the original automatic bias education. As a result, individuals did not receive the full lecture that participants in Study 1 did. These participants were also unable to complete some of the group-based demonstrations (e.g., the group-IAT). If these aspects are critical, then the bias education may not exceed the effectiveness of these other interventions.

However, whereas the Bias Blindspot Education (Pronin & Kugler, 2007) only describes automatic processes, and exposure to the IAT only provides an experiential component, the automatic bias incorporates both. Thus, I anticipated that the bias education would produce *stronger* effects than either of the other interventions.

2.3.1 Method

Participants

212 participants (96 male, mean age = 19.02) from the University of Virginia psychology participant pool participated for partial course credit.

Of these participants, 2 (1.0%) reported being American Indian/Alaska Native, 23 (10.8%) reported being East Asian, 19 (9.0%) reported being South Asian, 2 (1.0%)

reported being Native Hawaiian or other Pacific Islander, 13 (6.1%) Black or African American, 113 (62.7%) reported being White, 3 (1.5%) reported being more than one race – Black/White, 11 (5.2%) reported being more than one race – other, and 6 (2.8%) reported not knowing. 18 (8.5%) participants reported being of Hispanic or Latino ethnicity, 189 (89.2%) reported being not Hispanic or Latino origin, and 5 (2.3%) participants reported not knowing their ethnicity. Table 13 shows participants' responses to a brief demographic questionnaire regarding their current course enrollment.

Table 13. Participants' Course Demographics

Question	Frequency
Reason for Enrolling in the Course	
Elective	39
Major Requirement	22
Sounded Interesting	145
Only One that Would Fit My Schedule	6
Course Enrolled	
Introductory Psychology	68
Cognition	33
Personality	32
Social Psychology	42
Child Psychology	20
3000-level or above	1
Unreported	16
Are You Enjoying the Course?	
Yes	46
Planning on Being a Psychology Major?	
Definitely Yes	11
Probably Yes	16
Unsure	35
Probably No	49
Definitely No	101
Will You Enroll in More Psychology Courses?	
Definitely Yes	46
Probably Yes	54
Unsure	68
Probably No	32
Definitely No	12

Materials

Bias Education Manipulation. Participants viewed an edited, 10-minute clip of the bias education lecture from Study 1. This clip included particular points that I believed were important for understanding the key messages of the presentation and getting an experience with lack of insight into one's own mental processes. Specifically, participants viewed the parts of the talk demonstrating the McGurk effect (McGurk & MacDonald, 1976) and Shepherd's (1990) table illusion to illustrate that what is experienced in the world is not always the same as reality. Following that, participants completed a race-IAT individually on the laboratory computer.

Bias Blindspot Manipulation (Pronin & Kugler, 2007). Participants read a fictional *Science* article that reported actual research evidence suggesting that individuals are not always aware of the influences on their thoughts and behaviors. To demonstrate that automaticity can influence behavior, the article described several psychological studies showing how changes in environmental contexts can influence how people act in different social situations. For the complete article refer to Appendix G.

Bias Blindspot Questionnaire (Pronin & Kugler, 2007). The bias blindspot questionnaire asks participants to rate their own susceptibility, relative to that of their peers, to various biases on a 7-point Likert Scale (1=much less than the average student, 7=much more than the average student). For a full description, refer to Appendix H.

Beliefs about Bias Questionnaire. Participants completed the same questionnaire as described in Study 2.

Behavioral Intention and Policy Support Questionnaire. Participants reported how strongly they endorsed various scenarios relating to educating others about bias on a 5-point Likert scale (1=definitely no, 5=definitely yes). These scenarios ranged from implementing bias education (e.g., “Do you believe the University should consider adding an orientation program where first year students listen to a presentation designed to educate individuals about biases?”, “If the University added an orientation program where first year students listen to a presentation designed to educate individuals about biases, should it be mandatory?”, “Would you support implementing a program requiring students to attend a lecture intended to educate them about bias?”, “Would you support implementing a program requiring students to attend a lecture intended to educate them about bias if it meant you would be required to attend?”) to requesting time commitments from participants regarding bias education (e.g., “Would you be willing to be contacted by email with a handout designed to provide strategies for combating the adverse effects of bias?”, “Would you be willing to volunteer time to help plan a program intended to teach the UVA community about social biases (e.g., distributing flyers about the program, setting-up)?”).

If automatic bias education is effective, then I would expect that participants who receive it would be (1) more willing to devote time to helping with educational bias programs and (2) show higher agreement with policies that require others to receive bias education than participants who do not receive it.

Criminal Task (Park & Banaji, 2000). Participants completed the same task as described in Study 2.

Implicit Association Test (IAT; Greenwald, et al. 1998). The racial attitude IAT followed the same procedure as described in Study 2.

Procedure

After signing informed consent and completing demographics questions, participants were randomly assigned to receive one of the four conditions: a shortened automatic bias education, the bias blindspot education (Pronin and Kuger, 2007), experiential experience with a race-IAT only, or a control condition where they did not receive any bias education. Afterwards, all participants completed the questionnaires, the behavioral intention measure, and criminal task. The order of the dependent measures was completed in two blocks: the first block included a randomized order of the questionnaires, while the second included a fixed order of the criminal task (first) and the behavioral intention questionnaire (second). Finally, I randomized the order of the blocks to test whether the first block of dependent measures influenced participants' responses during the second block.

2.3.2 Results and Discussion

Results

For all dependent measures, an ANOVA was used first to test whether there were sampling effects that might have elicited differences in beliefs about bias. The following variables were included in the model: which class they were enrolled in, which psychology instructor they had, whether they were enjoying the course, and their reported reason (i.e., major versus elective) for enrolling in the course. Results revealed no significant differences for the any of the measures on their overall BaB score (Reason for Enrolling in the Course, $p = 0.08$, Course Enrolled In, $p = 0.91$, Enjoying the Course, $p =$

0.09, Planning on Becoming a Psychology Major, $p = 0.67$, Planning on Enrolling in More Psychology Courses, $p = 0.25$). Results also revealed no significant differences on any of the subscales. Likewise, results showed no significant differences on participants responses on the aggregated bias blindspot questionnaire (Reason for Enrolling in the Course, $p = 0.88$, Course Enrolled In, $p = 0.25$, Enjoying the Course, $p = 0.29$, Planning on Becoming a Psychology Major, $p = 0.68$, Planning on Enrolling in More Psychology Courses, $p = 0.24$). Results revealed two significant covariates for participants' responses regarding their behavioral intentions and policy support: Enjoying the Course ($p = 0.04$) and Reason for Enrolling in the Course ($p = 0.05$). The remaining covariates did not show significant differences (Course Enrolled In, $p = 0.12$, Planning on Becoming a Psychology Major, $p = 0.21$, Planning on Enrolling in More Psychology Courses, $p = 0.98$). However, including these two significant covariates (Enjoying the Course, Reason for Enrolling in the Course) in the full analyses of participants' behavioral intentions and policy support did not change the results. Therefore, these measures were not included in any of the analyses described below.

For the remaining analyses, an ANOVA was first used to examine differences across all conditions with post-hoc tests as the direct comparisons of conditions. Because the main focus was on how the bias education intervention compared to others, I also conducted two contrast analyses. For the first contrast, I compared the bias education treatment condition to the other conditions (i.e., IAT, Bias Blindspot, and control) while the second contrast compared all the interventions to the control condition to test whether, as a group, the interventions had an effect on the dependent variables.

For a complete list of means for all the below analyses, refer to Table 14.

Table 14. Participants' Means by Condition

	Automatic Bias Education		Experience with the IAT		Bias Blindspot Manipulation		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Beliefs about Bias								
Overall BaB Score	4.14	0.56	3.93	0.54	4.25	0.51	4.06	0.42
I Have Bias	4.72	0.96	4.58	0.91	4.93	0.94	4.77	0.68
Automatic Bias Influences Me	4.51	0.82	4.33	0.78	4.59	0.79	4.38	0.74
Self-Perception of Objectivity	4.79	1.16	4.85	0.84	4.43	1.00	4.70	1.15
Motivation to Address Automatic Bias	4.14	0.58	3.88	0.63	3.86	0.61	3.98	0.53
Knowing and Preventing Automatic Bias	3.15	1.06	3.48	0.96	2.93	0.99	3.46	0.93
Bias Blindspot Questionnaire	3.80	0.73	3.97	0.73	3.89	0.55	3.69	0.70
Criminal Task								
White Sounding Names	0.25	0.17	0.28	0.17	0.30	0.19	0.30	0.18
Black Sounding Names	0.34	0.25	0.37	0.26	0.38	0.24	0.30	0.24
Policy Support and Behavioral Intentions	4.82	0.82	4.72	0.88	4.91	0.84	4.93	0.89

Are there Differences in Participants' Beliefs about Bias?

Overall BaB Score. An ANOVA revealed a significant effect of condition on participants' overall *Beliefs about Bias* (BaB) score, $F(3, 208) = 3.52, p = 0.02, d = 0.26$. Post-hoc tests revealed that participants who completed the IAT-only intervention reported significantly smaller BaB ($M = 3.93, SD = 0.54$) than participants who received

the Bias Blindspot intervention ($M = 4.25$, $SD = 0.51$) but not significantly different from bias education participants ($M = 4.14$, $SD = 0.56$) or control participants ($M = 4.06$, $SD = 0.42$). In the follow-up contrast analyses, the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 208) = 0.48$, $p = 0.49$, $d = 0.10$, nor did the bias interventions show significantly different BaB scores than the control condition, $F(1, 208) = 0.38$, $p = 0.54$, $d = 0.09$. Next, I examined whether there were differences across interventions on any of the specific subscales of BaB.

I Have Bias. In contrast to my hypothesis, an ANOVA revealed no significant effect of condition, $F(3, 208) = 1.42$, $p = 0.24$, $d = 0.17$. A contrast revealed the bias education did not produce higher beliefs that *I Have Bias* compared to the other conditions, $F(1, 208) = 1.42$, $p = 0.24$, $d = 0.16$. Likewise, contrast coding showed that there was no significant differences between receiving an intervention and the control, $F(1, 208) = 0.04$, $p = 0.84$, $d = 0.03$.

Automatic Bias Influences Me. An ANOVA revealed no significant effect of condition, $F(3, 208) = 1.23$, $p = 0.30$, $d = 0.15$. A contrast revealed the bias education did not produce higher beliefs that *Automatic Bias Influences Me* compared to the other conditions, $F(1, 208) = 0.37$, $p = 0.54$, $d = 0.08$. Likewise, contrast coding showed that there was no significant difference between participants who received an intervention and control participants, $F(1, 208) = 0.61$, $p = 0.44$, $d = 0.11$.

Self-Perception of Objectivity. In contrast to my hypothesis, an ANOVA revealed no significant effect of condition, $F(3, 208) = 1.65$, $p = 0.18$, $d = 0.18$. Planned contrasts

revealed the bias education did not produce higher *Self-Perception of Objectivity* compared to the other conditions, $F(1, 208) = 0.61, p = 0.44, d = 0.11$. Results also showed that there was no significant difference between receiving a bias intervention and the control condition, $F(1, 208) = 0.01, p = 0.94, d = 0.01$.

Motivation to Address Automatic Bias. An ANOVA revealed marginal significant differences based on condition, $F(3, 208) = 2.52, p = 0.06, d = 0.22$. In line with the hypothesis, a contrast revealed that the bias education produced more agreement with *Motivation to Address Automatic Bias* compared to the other conditions, $F(1, 208) = 6.34, p = 0.01, d = 0.35$. However, contrast coding showed that there was no significant difference between receiving an intervention and the control, $F(1, 208) = 0.05, p = 0.83, d = 0.03$.

Knowing and Preventing Automatic Bias. An ANOVA revealed a significant effect of condition, $F(3, 208) = 3.79, p = 0.01, d = 0.27$. Post-hoc tests revealed that participants who received the Bias Blindspot intervention ($M = 2.93, SD = 0.99$) were significantly less likely to believe that they knew and could prevent automatic biases than participants who only completed the IAT as the intervention ($M = 3.48, SD = 0.96$) or control participants ($M = 3.45, SD = 0.93$) while there was no significant difference compared to those who received the bias education ($M = 3.15, SD = 1.06$). Planned contrasts revealed the bias education did not produce higher beliefs in their *Knowing and Preventing Automatic Bias* compared to the other conditions, $F(1, 210) = 0.79, p = 0.38, d = 0.12$. Similarly, contrast analyses showed that there was a marginally significant

difference between participants receiving a bias intervention and control participants, $F(1, 208) = 3.04, p = 0.08, d = 0.24$.

The Type of Intervention Received did not affect Participants' Responses on the Bias Blindspot Questionnaire

To test for differences on participants' responses to the Bias Blindspot questionnaire, responses to the individual questions were aggregated with higher scores indicating participants' increased belief that they could be susceptible to automatic processes. Overall, an ANOVA revealed no significant differences based on condition, $F(3, 208) = 1.64, p = 0.18, d = 0.18$. A contrast revealed the bias education did not produce higher beliefs compared to the other conditions, $F(1, 208) = 0.26, p = 0.61, d = 0.07$. Likewise, planned contrasts showed marginal significant differences between the interventions and the control, $F(1, 208) = 3.34, p = 0.07, d = 0.25$.

Learning about Bias did not Influence Participants' Judgments on the Criminal Task

A 2 (Target Race: Black versus White) X 4 (Condition: Bias Education versus IAT versus Bias Blindspot versus Control) ANOVA was conducted on the proportions of names identified as criminals. Across conditions, Black sounding names were more likely than White sounding names to be misjudged as criminals, $F(1, 208) = 22.94, p < 0.0001, d = 0.33$. Results also showed a significant interaction of Target Race X Condition for participants' criminal judgments of Black and White sounding names, $F(3, 208) = 2.73, p = 0.05, d = 0.23$, indicating that the differential tendency to misjudge Black and White sounding names was less for some conditions than others. Surprisingly, control participants showed the smallest judgment difference: they were equally likely to judge

Black sounding names ($M = 0.30$, $SD = 0.24$) and White sounding names ($M = 0.30$, $SD = 0.17$) as criminals.

In addition, planned contrasts revealed that the bias education did not result in lower judgments of criminal for Black sounding names, $F(1, 208) = 0.12$, $p = 0.73$, $d = 0.05$, or White sounding names, $F(1, 208) = 2.50$, $p = 0.12$, $d = 0.22$, compared to the other conditions. Likewise, contrast analyses showed the other interventions did not show lower judgments of criminal for Black sounding names, $F(1, 208) = 2.79$, $p = 0.10$, $d = 0.23$, or White sounding names, $F(1, 208) = 0.59$, $p = 0.44$, $d = 0.11$, compared to control participants.

There were no Differences on Participants' Reported Behavioral Intentions or Support for Policies Involving Bias Education for Others

An ANOVA revealed no significant effect of condition, $F(3, 208) = 0.66$, $p = 0.57$, $d = 0.11$. A contrast revealed the bias education did not result in greater behavioral intentions or policy support compared to the other conditions, $F(1, 208) = 0.05$, $p = 0.82$, $d = 0.03$. There was also no significant difference between the interventions and the control for participants' reported behaviors, $F(1, 208) = 0.72$, $p = 0.40$, $d = 0.12$.

Discussion

Study 3 examined whether the *experiential* component of the bias education impacted beliefs and social judgment beyond other successful interventions that educate individuals about bias. To test, I compared a condensed version of the bias education with two other effective interventions: (1) Pronin and Kugler's (2007) bias blindspot exercise, and (2) simple exposure to the IAT without accompanying experiential education.

Learning about Bias did not Influence Participants Beliefs about Bias

Overall, the results suggested that all three interventions were ineffective compared to the control condition. The small number of reliable effects do not clearly favor one intervention over the others. For example, while participants who received the bias blindspot intervention reported significantly greater overall BaB scores compared to those participants who completed the IAT as the intervention, participants who received automatic bias education showed marginal higher agreement in *Motivation to Address Automatic Bias* compared to the other interventions. Similarly, results showed no significant differences for participants' responses on the bias blindspot questionnaire as a function of the bias intervention received. Because of the lack of consistency and the small number of significant effects amid a large number of inferential tests, my conservative conclusion is that all three interventions were ineffective at changing beliefs about bias, and behavioral judgments and policy support linked to bias.

2.4 Studies 4-6: Focused Studies Examining the Shortened Bias Education

Overview of Studies 4-6

Studies 1 and 2 showed a clear influence of the educational intervention on beliefs about bias, and it even appeared to provide additional influence beyond naturalistic changes in students' beliefs about bias – perhaps as a consequence of the “college experience.” In Study 3, I tried to distill the essentials from the education into a short video so that I could examine the intervention experimentally in the laboratory. The lack of effects in Study 3 suggest two possible explanations: (1) the video version may have lost critical features for making the intervention effective, or (2) the student sample –

recruited in the Spring semester – may have already changed on their beliefs about bias as a function of their college experience making it more difficult to detect further change in those beliefs. Therefore, I conducted three follow-up studies to demonstrate an effect of the intervention, or distinguish between these possibilities for it failing to have an impact. To conserve space, I report all three studies together.

Study 4. I first addressed the possibility that the college experience fosters beliefs in the existence of bias (Pascarella et al., 1996), and reduces the effectiveness of the education intervention later in the participants' college careers. There are two aspects of the college experience that are important to assess: those who report having any college experience (past or present), and those who report being currently enrolled in college. The former would suggest that it is something about the college experience itself, and that the effects are long lasting. On the other hand, finding differences between those who are currently enrolled versus those who are not would suggest that it is something about being immersed in a college environment that makes people more open to the idea of the existence of bias. This would also suggest that these effects would be fleeting and would dissipate over time once the individual is no longer enrolled. A final possibility is that completion of a psychology course might make individuals more open to the existence of bias.

To test this idea, I conducted a study with the identical materials and procedure on an additional sample recruited through Amazon's Mechanical Turk (Mturk) online participant pool to compare (1) people who attended college versus those who did not and (2) those who reported being currently enrolled in college versus those who are not. The

Mturk site attracts a heterogeneous sample by age and education. Moreover, because Mturk participants are paid for their time, I was able to use the identical materials from Study 3 without needing to condense the manipulations and/or dependent measures (this would not have been possible using Project Implicit as studies run through its infrastructure are limited to 10-15 minutes because participants are volunteers and are not compensated).

Studies 5 and 6. The results to Study 4 showed the same pattern of results as Study 3 for participants' reported *beliefs about bias* – participants receiving the bias education did not differ from control participants. Therefore, for Studies 5 and 6 I returned to the Project Implicit infrastructure and pursued high-powered tests to fully consider the possibility that the shortened, video-based bias education is simply not impactful enough to be an effective intervention.

In Study 5 participants watched the same condensed bias education video used in Study 3 and 4. Participants who received the education reported stronger *beliefs about bias* than participants who had not received any education. However, there was a large drop-out rate in the bias education condition compared to the control condition (29.8% versus 7.3%). It is possible that the results were due to differential attrition rates – perhaps those most skeptical about automatic bias were most likely to leave the study during the video.

Thus in Study 6, I wanted to replicate the result while trying to minimize differential attrition. I also assessed whether the drop-out rates in Study 5 were symptomatic of the content participants were asked to watch or whether it occurred

simply because participants were being asked to watch a video. To test, I used two other popular videos as control conditions: a clip from the “Last Lecture” (a famous lecture by computer scientist Randy Pausch) and a “Matt” dancing video (Matt films himself dancing with people from around the world). The “Last Lecture” video was chosen because it was a similar length and lecture format as the intervention. The dancing video was also a video, but a substantially different format and content. The results showed much weaker evidence that the shortened bias education changed *beliefs about bias*. Further, the shortened bias education continued to produce larger drop-out rates (23.3%) compared to either the “Last Lecture” or dancing video (14.2% and 9.9%, respectively). So, I cannot rule out the possibility of selective attrition being responsible for the observed differences between conditions. Conservatively, I conclude that the bias education may require other features of the in-person education event to change beliefs about bias.

2.4.1 Method

Participants

Study 4. 452 participants (167 male, 8 unreported, mean age = 31.22) from Amazon Mechanical Turk website participated for \$0.25 payment.

Of these participants, 12 (2.7%) reported being American Indian/Alaska Native, 13 (2.9%) reported being East Asian, 11 (2.4%) reported being South Asian, 2 (0.4%) reported being Native Hawaiian or other Pacific Islander, 29 (6.4%) Black or African American, 349 (77.2%) reported being White, 4 (1.0%) reported being more than one race – Black/White, 17 (3.8%) reported being more than one race – other, 7 (1.5%)

reported not knowing, and 8 (1.8%) chose not to answer. 32 (7.1%) participants reported being of Hispanic or Latino ethnicity, 379 (83.8%) reported being not Hispanic or Latino origin, 24 (5.3%) participants reported not knowing their ethnicity, and 17 (3.8%) chose not to respond.

389 (86.1%) of the sample reported having some college experience. 81 (18.1%) of participants reported being currently enrolled as a student. 323 (71.5%) of participants reported having taken at least one psychology course.

Study 5. 372 individuals (130 male, 1 unreported, mean age = 29.96) visiting the Project Implicit research website (<https://implicit.harvard.edu/>) volunteered to complete this study.

Of the participants 2 (1.0%) reported being American Indian/Alaska Native, 14 (3.8%) reported being East Asian, 12 (3.2%) reported being South Asian, 5 (1.3%) reported being Native Hawaiian or other Pacific Islander, 27 (7.3%) reported being Black or African American, 257 (69.1%) reported being White, 5 (1.3%) being biracial – Black/White, 27 (7.3%) reported being more than one race – other, 20 (5.4%) reported not knowing, and 3 (0.3%) chose not to respond. 50 (13.4%) participants reporting being of Hispanic or Latin descent, 277 (74.5%) participants reported not being of Hispanic or Latin descent, 19 (5.1%) participants reported not knowing and 26 (7.0%) chose not to respond.

Study 6. 759 individuals (108 male, 2 unreported, mean age = 30.79) visiting the Project Implicit research website (<https://implicit.harvard.edu/>) volunteered to complete this study.

Of the participants 5 (1.0%) reported being American Indian/Alaska Native, 8 (1.1%) reported being East Asian, 11 (1.4%) reported being South Asian, 2 (0.3%) reported being Native Hawaiian or other Pacific Islander, 80 (10.5%) reported being Black or African American, 561 (73.9%) reported being White, 9 (1.2%) being biracial – Black/White, 50 (6.6%) reported being more than one race – other, 27 (3.6%) reported not knowing, and 6 (0.8%) chose not to respond. 74 (9.7%) participants reporting being of Hispanic or Latin descent, 575 (75.6%) participants reported not being of Hispanic or Latin descent, 37 (4.9%) participants reported not knowing and 73 (9.6%) chose not to respond.

New Materials

Last Lecture Video. This 10-minute abbreviated lecture (<http://youtu.be/ncoSRKoU6GQ>) is given by Dr. Randy Pausch, a terminally ill professor, during an episode of Oprah. Titled “The Last Lecture,” Dr. Pausch offers his audience life lessons he believes are particularly important for achieving one’s childhood dreams.

Dancing Video. This 5-minute documentary (<http://youtu.be/zlfKdbWwruY>) is of a man who travels to film himself dancing with people from around the world.

Procedure

Study 4. Participants completed the identical procedure as reported in Study 3 including the other bias intervention conditions and dependent variables.

Study 5. About half the participants were randomly assigned to watch the bias education video and then complete the race-IAT, the beliefs about bias questionnaire, and

self-reported racial attitude measures. This presentation order was fixed. The other participants completed the same procedure except that there was no video.

Study 6. About half the participants were randomly assigned to watch the bias education video, followed by the race-IAT, bias questionnaire, and self-reported racial attitudes in a fixed order. The other participants were randomly assigned to watch either the dancing video or the “Last Lecture” video followed by the same measures in the same fixed order.

2.4.2 Results and Discussion

Results

Study 4

An ANOVA was first used to examine differences across all conditions with post-hoc tests as the direct comparisons of conditions. Because the main focus was on how the bias education intervention compared to others, I also conducted two contrast analyses. For the first contrast, I compared the bias education treatment condition to the other conditions (i.e., IAT only, Bias Blindspot, and control) while the second contrast compared all the interventions to the control condition to test whether, as a group, the interventions had an effect on the dependent variables.

Results revealed a marginally significant effect of condition on participants' overall *Beliefs about Bias* (BaB) score, $F(3, 445) = 2.53, p = 0.06, d = 0.15$. In the follow-up contrast analyses, the bias education treatment condition was not different from the other conditions (i.e., IAT only, Bias Blindspot, and control), $F(1, 447) = 1.90, p = 0.17, d = 0.13$, nor did the bias interventions show significantly different BaB scores than

the control condition, $F(1, 447) = 1.95, p = 0.16, d = 0.13$. These contrast effects suggest that the marginally significant effect reflect effectiveness of neither the bias education treatment nor the treatments in general compared to the control condition.

I next repeated the ANOVA and two contrasts for the individual subscales. Results revealed no reliable effects that could be attributed to the bias education interventions, or the interventions in general. I further tested whether having college experience moderated participants' *beliefs about bias*. It did not (all p 's ≥ 0.30). Another possibility is that it is not the college experience itself, but rather being immersed in a college environment that changes participants' *beliefs about bias*. Thus, I compared those who were currently enrolled as students to those who were not. An ANOVA revealed that it did not moderate the participants' *beliefs about bias* (all p 's ≥ 0.17).⁸

The remainder of the results showed similar findings to the ones observed in Study 3: the interventions did not differentially impact the other dependent measures. Complete analyses of the other dependent variables and subscales appear in the Supplement.

Study 5

Because the focus was on examining participants' *beliefs about bias* the other dependent measures (race-IAT and self-reported racial attitudes) are not described here. No differences were observed on those dependent variables across conditions. Complete analyses appear in the Supplement.

---

⁸ An alternative consideration is that having knowledge of psychology material (measured by completing psychology courses) moderated participants' *beliefs about bias*. It did not (all p 's ≥ 0.14).

Participants who watched the bias education video reported higher overall BaB scores ($M = 4.18$, $SD = 0.62$) than control participants did ($M = 3.93$, $SD = 0.54$), $t(368) = 4.21$, $p < 0.0001$, $d = 0.44$. Participants who received the bias education also reported significantly more agreement on *I Have Bias* ($M = 4.74$, $SD = 0.96$) than control participants ($M = 4.47$, $SD = 0.97$), $t(368) = 2.58$, $p = 0.01$, $d = 0.27$. Likewise, participants who received the bias education more strongly agreed that *Automatic Bias Influences Me* ($M = 4.61$, $SD = 0.87$) than participants who did not, ($M = 4.20$, $SD = 0.90$), $t(368) = 4.31$, $p < 0.001$, $d = 0.45$). Results also revealed that bias educated participants showed marginally more agreement in *Motivation to Address Automatic Bias* ($M = 4.26$, $SD = 0.62$) than participants who did not ($M = 4.15$, $SD = 0.52$), $t(280) = 1.80$, $p = 0.07$, $d = 0.22$.

Similarly, participants who received the bias education showed marginally less agreement on *Self-Perception of Objectivity* ($M = 4.70$, $SD = 0.99$) than control participants, ($M = 4.89$, $SD = 1.06$), $t(367) = 1.76$, $p = 0.08$, $d = 0.18$. Finally, participants who viewed the bias education video showed significantly less agreement in *Knowing and Preventing Automatic Bias* ($M = 3.39$, $SD = 1.06$) than control participants, ($M = 3.69$, $SD = 1.06$), $t(367) = 2.65$, $p = 0.01$, $d = 0.28$.

Differential Drop-Out Rates. The findings suggests that the intervention was effective compared to a “no video” control condition. However, there was a large differential drop-out rate for the participants in the bias education condition compared to the control condition (29.8% versus 7.3%). Thus, it was important to examine whether there were any demographic differences between those that dropped out versus those who

did not. To test, I first used a chi-square test to compare the demographics of those who had versus had not completed the study.⁹ The results revealed there were no demographic differences on their reported sex ($p = 0.57$), race ($p = 0.17$), ethnicity ($p = 0.45$), age ($p = 0.25$), political identity ($p = 0.59$), religious identity ($p = 0.79$), or education level ($p = 0.07$).

I next used a logistic regression to test whether the drop-out rates varied differentially as a function of the condition participants were assigned, using participants' demographics and condition as the independent variables and participants' completion status as the dependent variable. The results showed no demographic differences between those participants who completed the study and those who did for sex ($p = 0.84$), race ($p = 0.79$), ethnicity ($p = 0.79$), age ($p = 0.49$), political identity ($p = 0.37$), religious identity ($p = 0.40$), or education level ($p = 0.42$). Despite not finding any demographic differences, it is possible that I may not have measured the factors that could show how the differential attrition (rather than the manipulation) affected the outcome. To examine this possibility, Study 6 attempted to increase the similarity across conditions as a means of reducing the differential attrition.

Study 6

The analyses focused on examining participants' *beliefs about bias*; therefore, the other dependent measures (race-IAT and self-reported racial attitudes) are not described here. No differences were observed across conditions. Complete analyses appear in the Supplement.

---

⁹ The exception to this test was the age variable. For this, I used a logistic regression with age as the predictor and participants' completion status as the dependent variable.

Differential Drop-Out Rates. Because Study 5 revealed high drop-out rates for the experimental condition, I first examined the drop-out rates for each condition. Results revealed a 23.3% drop-out rate for participants who received the shortened bias education. In contrast, participants who were assigned to view the “Last Lecture” and dancing video had a 14.2% and 9.9% drop-out rate, respectively. My alternative manipulations reduced, but did not eliminate differential drop-out. This suggests that part of the differential attrition is probably due to the presentation of a long video, but part is also likely a consequence of the content of the video itself. The latter offers the most risk for interpreting the observed effects as being due to the intended experimental manipulation.

In order to see if participants who completed the study were different from those who did not, I first used a chi-square test to compare the demographics of those who had versus had not completed the study.¹⁰ Overall, results showed that the participants were not significantly different from each other on any of the reported demographics including: sex ($p = 0.22$), race ($p = 0.09$), ethnicity ($p = 0.63$), age ($p = 0.14$), political identity ($p = 0.77$), religious identity ($p = 0.69$), or education level ($p = 0.68$). I next used a logistic regression to test whether the drop-out rates varied differentially as a function of the condition participants were assigned, using participants’ demographics and condition as the independent variables and participants’ completion status as the dependent variable. Results revealed no demographic differences between conditions for those participants who completed the study for sex ($p = 0.39$), race ($p = 0.92$), ethnicity

---

¹⁰ The exception to this test was the age variable. For this, I used a logistic regression with age as the predictor and participants’ completion status as the dependent variable.

($p = 0.49$), age ($p = 0.37$), political identity ($p = 0.81$), religious identity ($p = 0.16$), or education level ($p = 0.96$). As in Study 5, the lack of differences between completers and drop-out participants increases confidence somewhat that there were no systematic factors. However, the tested variables were demographics, not psychological variables such as pre-existing beliefs about bias. Therefore, the following results should be interpreted with caution.

The Effect of the Bias Education Intervention on Beliefs about Bias. For all remaining analyses, an ANOVA was first used to examine differences across all conditions with post-hoc tests as the direct comparisons of conditions. Because the main focus was on how the bias education intervention compared to the control videos, I also conducted a contrast comparing the shortened bias education treatment to the control conditions (i.e., the “Last Lecture” and dancing video). For a complete list of means for all the below analyses, refer to Table 15 below.

Table 15. Means for the Between-Groups Comparisons

	Bias Education		“The Last Lecture”		Dancing Video	
	Mean	St Dev	Mean	St Dev	Mean	St Dev
Overall BaB Score	4.11	0.63	4.02	0.57	4.07	0.60
I Have Bias	4.66	0.98	4.61	0.98	4.57	0.96
Automatic Bias Influences Me	4.52	0.95	4.30	0.98	4.39	0.91
Self-Perception of Objectivity	4.70	1.12	4.72	1.10	4.81	1.10
Motivation to Address Automatic Bias	4.18	0.66	4.15	0.60	4.14	0.61
Knowing and Preventing Automatic Bias	3.38	1.13	3.63	1.10	3.32	1.18

Overall BaB Scores. Results revealed that participants who watched the shortened bias education video reported similar overall BaB scores ($M = 4.11$, $SD = 0.63$) than

either the participants who watched “The Last Lecture” clip ($M = 4.02$, $SD = 0.57$) or the dancing clip ($M = 4.07$, $SD = 0.60$), $F(2, 746) = 1.13$, $p = 0.32$, $d = 0.08$. Planned contrasts also revealed the bias education did not produce higher overall BaB scores compared to the control conditions, $F(1, 747) = 1.70$, $p = 0.19$, $d = 0.10$.

I Have Bias. Results revealed no significant differences based on conditions, $F(2, 745) = 0.57$, $p = 0.56$, $d = 0.06$. The planned contrast revealed that the shortened bias education did not make participants more likely to agree on *I Have Bias* than participants who were assigned the control condition, $F(1, 746) = 0.93$, $p = 0.33$, $d = 0.07$.

Automatic Bias Influences Me. An ANOVA revealed a significant effect of condition, $F(2, 745) = 3.68$, $p = 0.03$, $d = 0.14$. Post-hoc tests revealed that participants who received the bias education showed significantly more agreement on *Automatic Bias Influences Me* ($M = 4.52$, $SD = 0.95$) than participants who watched “The Last Lecture” ($M = 4.30$, $SD = 0.98$), while there was no difference between participants who watched the bias education or who watched the dancing video ($M = 4.38$, $SD = 0.91$). Post-hoc tests also revealed that participants who watched “The Last Lecture” showed similar beliefs on *Automatic Bias Influences Me* compared to participants who watched the dancing video.

A planned comparison revealed that the participants who viewed the shortened bias education showed significantly more agreement on *Automatic Bias Influences Me* compared to the control conditions, $F(1, 748) = 6.75$, $p = 0.01$, $d = 0.19$.

Self-Perception of Objectivity. Results showed that there were no significant differences based on conditions, $F(2, 744) = 0.65$, $p = 0.52$, $d = 0.06$. Planned contrasts

revealed that the participants who received the shortened bias education showed similar beliefs on *Self-Perception of Objectivity* to control participants, $F(1, 745) = 0.62, p = 0.43, d = 0.06$.

Motivation to Address Automatic Bias. An ANOVA revealed that there were no significant differences based on conditions, $F(2, 745) = 0.36, p = 0.69, d = 0.06$. Likewise, a planned contrast revealed that the participants who received the shortened bias education did not significantly differ on *Motivation to Address Automatic Bias* compared to control participants, $F(1, 746) = 0.68, p = 0.41, d = 0.06$.

Knowing and Preventing Automatic Bias. Results revealed a significant effect of condition, $F(2, 745) = 4.37, p = 0.01, d = 0.15$. Post-hoc tests showed that participants who viewed “The Last Lecture” video reported significantly higher agreement on *Knowing and Preventing Automatic Bias* ($M = 3.63, SD = 1.10$) than either participants who viewed the dancing video ($M = 3.32, SD = 1.18$) or the shortened bias video ($M = 3.38, SD = 1.13$). Post-hoc tests revealed no significant differences between participants who viewed the bias education or the dancing video.

A planned contrast revealed that the participants who received the shortened bias education did not significantly differ on *Knowing and Preventing Automatic Bias* compared to control participants, $F(1, 746) = 1.24, p = 0.27, d = 0.08$.

Discussion

Studies 4-6 focused on examining whether the shortened automatic bias education was an effective intervention. Study 4 showed no effect of the intervention in a different sample, and no evidence that college experience was a moderating influence on the effect

of the intervention. Thus, Study 5 and 6 were designed to provide a more rigorous test of the impact of the shortened bias education. The results from the two studies were inconsistent: Study 5 showed that the shortened automatic bias education was effective, while the results from Study 6 suggest that the shortened bias education was not clearly better at producing change in people's *beliefs about bias* than the control condition.

Importantly, the results from Study 5 and 6 revealed large attrition rates, raising the possibility that the findings could also be due to selective drop-out. Although the results did not reveal any demographic differences between the participants who completed the study and those who did not, the possibility that the drop-out pattern spuriously produced the results remains. Given the lack of consistency, coupled with the attrition problem, my conservative conclusion is that the shortened automatic bias education was not effective enough to significantly change participants' *beliefs about bias* and may not be a viable intervention to pursue. The fact that there was differential attrition itself is an indicator that the shortened, video-based version of the intervention was ineffective – people were much more likely to just leave the study compared to watching other videos. An effective intervention needs to engage viewer's interest so that they attend, learn, and internalize the educational messages.

Chapter 3

3.1 Summary of Studies

This dissertation has contributed to scientific knowledge by providing insight in how diversity training might be altered to increase their effectiveness for its participants. Using Wilson and Brekke's mental contamination model (1994), the goal of the present

work was to investigate whether providing experiential education about one's own biases was enough to shift beliefs about bias and social behavior. In Study 1 and 2, I examined the immediate and long-term impact on *beliefs about bias* of an interactive lecture about how biases are an ordinary part of human thinking. I found that automatic bias education was effective: individuals completing the bias questionnaire after receiving the education were more likely to show stronger *beliefs about bias* (Study 1) and that this single 1-hour lecture produced long-term durability for participants' *beliefs about bias* (Study 2). However, I also found that the automatic bias education did not produce changes in participants' automatic and self-reported racial attitudes (Study 2), nor did it change participants' judgments in a stereotyping task (Study 2 & 3).

In Studies 4-6, I used a shortened video version of the automatic bias education to provide a more focused investigation of its effectiveness. The results from these studies revealed some between-group differences. However, these differences only occurred when there was differential drop-out in the intervention condition versus the control conditions. The fact that I did not find consistent effects of the intervention across studies, coupled with the differential attrition rate, leads me to believe that the effects are spurious and the shortened intervention is not effective.

3.2 Why the Shortened Bias Education was not as Impactful as the Original Education

There are several reasons why the shortened version of the automatic bias education may not have been as effective as the original. One possibility is that the shortened bias education did not capture the original content that was necessary to impact

individuals' *beliefs about bias*. The original automatic bias education was about an hour in length; in comparison, the shortened one was distilled to 10 minutes. It is possible that the material that I left out was vital to the education. For example, the original education included two IATs, as well as an explanation of what an automatic association is, how it is measured, and why procedural factors, such as the order of the pairing condition on the IAT (e.g., Black people paired with unpleasant words first versus Black people paired with pleasant words first) is not an alternative explanation for the results. Completing a non-threatening IAT first (Virginia Tech compared to UVA), as well as learning the additional information about the IAT may have made participants less skeptical about the validity of the test, making them more likely to accept that they have automatic bias. Likewise, in the original version participants are shown how White and Black people score, on average, on the race-IAT. This may help strengthen the argument that everyone, not just them, has bias. Finally, participants received four demonstrations about unconscious or automatic processes; in the distilled version, participants only viewed two. In this case, it was not that I did not include the information that was needed; rather, I did not include enough demonstrations that would clearly illustrate the point in order to make a compelling case.

Another possibility is that the automatic bias education needs to be experienced in person, surrounded by other participants. Research has shown that the presence of others can facilitate learning (see for example, Doise & Mugny, 1984) and lead to social comparisons (Kruglanski & Mayseless, 1990). Indeed, the presence of others has been shown to increase self-reflection, allowing members to assess discrepancies between their

ideal self and actual self (for a review see Levine & Resnick, 1993). These findings may explain why the shortened bias education was not as effective as the original. For example, the McGurk demonstration (McGurk & MacDonald, 1976) is a particularly impactful group demonstration for showing that speech perception uses both visual and auditory information. In it, participants are asked to identify the sound that a man produces during a video clip. Half of the participants are asked to do this while watching the video, while the others are asked to identify the sound without watching. When asked to report what they heard, they learn that the half of the room that listened with their eyes closed reported hearing “ba ba” while the half that watched and listened reported hearing “da da.” This discrepancy allows participants to experience, very directly and dramatically, that the same auditory information can be experienced differently depending on what they are seeing. Likewise, the group demonstration IAT provides strong experiential evidence that everyone has automatic bias. In this demonstration, participants complete a race-IAT aloud. This raises participants’ awareness that not only are they having trouble pairing good words with Black faces, but their neighbors are as well. Because the shortened version was experienced individually, it did not provide the information that others also experience the same difficulty during the race-IAT. Thus, both demonstrations may have been less effective in the distilled version than in the original; experiencing the demonstrations alone may fail to adequately show the discrepancy between people’s experience and reality, and show to participants that everyone has bias.

In addition, previous research (Heesacker Petty, & Cacioppo, 1983; Hovland & Weiss, 1951) demonstrated that individuals are more likely to be persuaded when they believe that the speaker is credible. In comparison to the original intervention, the shortened version did not introduce the speaker or his credentials to the audience. Thus, it is possible that the lecturer was not perceived as credible in the video as compared to the original, in person education. Likewise, research has shown that self-disclosure is an important part of trusting others (Kerr, Sattin, & Trost, 1999; Wheelless & Grotz, 1977). In the original version, the lecturer provided information regarding his own automatic racial biases whereas the shorter version provided no such information. It is possible that this additional information may also increase the speaker's credibility in the full length version.

In sum, the results suggest that an effective bias intervention may require experiential live-group demonstrations. However, one question that remains is how many demonstrations are necessary to make people aware of and motivate change in their biases.

3.3 Why Some Factors May be More Durable than Others

Study 2 provided evidence that the effects of automatic bias education were durable for some of the individual components of *beliefs about bias* (*I Have Bias* and *Automatic Bias Influences Me*). However, the remaining factors (*Self-Perception of Objectivity*, *Motivation to Address Automatic Bias*, *Knowing and Preventing Automatic Bias*) did not show the same durability.

There are several reasons why these latter factors may not be as durable. For example, the lack of durability of the factor *Self-Perception of Objectivity* may be due to people's unwillingness to admit they could be non-objective. Indeed, Pronin and Kugler (2007) show that individuals more easily recognize bias in others than they recognize in themselves. Likewise, Uhlmann and Cohen (2007) showed that participants strongly agreed that they were objective ($M = 8.05$, $SD = 1.18$ on a 10-point scale). Finally, several experiments have unsuccessfully tried to induce individuals into feeling *non-objective* (Nagel, Joy-Gaba & Nosek, 2010; Lindner, Nosek, & Graser, 2009). Thus, although Study 1 showed a moderate belief change ($d = 0.25$) in participants' *Self-Perception of Objectivity* as a result of receiving bias education, it could be that without persistent reminders and evidence, individuals revert back to believing they are objective. It is curious to note that people's beliefs that they were biased persisted when their beliefs in their lack of objectivity did not. This may be a consequence of it being very identity challenging to claim non-objectivity, whereas affirming the potential for bias may be less so. People may resist affirming they are non-objective because doing so suggests something about their identity or personality. On the other hand, affirming bias may only imply to participants that they are acknowledging they have bias on some occasions, not that they would act on it in a particular decision-making context or that it is a part of their identity or personality.

The factor *Motivation to Address Automatic Bias* showed the smallest initial effect compared to the other factors ($d = 0.14$), suggesting that it was the least impacted by the educational intervention. One possibility is that individuals' motivation dissipated

over time and, because it was a weak change to begin with, no change survived. Another possibility is due to the education itself. The majority of the education focused on educating individuals about automatic bias, what it is and what kinds of effects it can have for decision makers. There was much less time focused on teaching individuals ways in which they could combat its effects. As a result, participants may have left with a clear understanding of the pervasiveness of automatic bias but have been less confident, and thus motivated, on how to counteract the effects of bias. This may also explain why *Knowing and Preventing Automatic Bias* also did not show persisting change. The education effectively convinced people that bias occurs, but not what to do about it.

3.4 Automatic Bias Education Differentially Impact Social Groups

Research has shown that most diversity training programs focus on historical prejudices (Von Bergen, Soper, & Foster, 2002) and assume that bias is intentional (Jackson & Joshi, 2010; Kalev, Dobbin, & Kelly, 2006). In addition, many organizations have mandatory participation in diversity training (SHRM, 2010). Two potential problems arise with programs such as these: (1) participants may resent being required to attend and (2) participants – particularly white men who are more highly represented in senior positions in most organizations – may feel targeted and offended by the implication that they could be prejudiced. Moreover, by focusing on historical prejudices, these programs may fail to communicate the message that everyone has bias.

Study 1 showed that not only was automatic bias education effective at changing participants' *beliefs about* bias but it also demonstrated that some social groups were particularly impacted by the education. Specifically, females showed somewhat larger

differences than males from pre to post-test on the factors *Automatic Biases Influence Me* and *Self-Perception of Objectivity*. Likewise, Black participants showed some of the largest changes in their *beliefs about bias* from pre to post-test on the factors *I Have Bias*, *Automatic Bias Influences Me*, and *Self-Perception of Objectivity*. The fact that the bias education impacted not only the majority members, but also non-majority members is important; it provides evidence that the education is effective at showing that every person has bias.

3.5 Why Automatic Bias Education was Ineffective at Changing Participants' Attitudes and Judgments

An important aspect of diversity training is to help foster change so that its participants will be less likely to engage in biased judgments and behaviors in future situations. The current bias education was unsuccessful in this pursuit: Study 2 revealed that automatic bias education did not produce changes in participants' automatic and self-reported racial attitudes. Likewise, Study 2 and 3 revealed that it did not change participants' judgments in a stereotyping task.

In Study 2, participants' attitudes and judgments were measured months after the intervention initially occurred. One possibility is that the intervention's impact dissipates over time. Thus, one question that remains is whether the original automatic bias education can reduce bias in judgments in the short-term. Another possibility is that the education did not incorporate enough of the "right" material. Given that research has shown that individuals are more likely to be motivated when they believe the effort will be successful (Weiner, 1985), it is possible that individuals left the education with

uncertainty of how they might minimize the influences of automatic bias (as shown by the lack of durability on *Motivation to Address Automatic Bias* and *Knowing and Preventing Automatic Bias*). As a result, they may not have been as motivated to closely examine the direction and magnitude of their biases (step 2 and 3 in Wilson and Brekke's mental contamination model). In turn, they would have been unable to change their biased responses in future decision-making situations (step 4).

The results from Studies 4-6 suggest that the manipulation used (i.e., the shortened version of the bias education) was ineffective at changing participants' *beliefs about bias*. Thus if the intervention is unable to produce change in people's beliefs, it should follow that it should not produce change in judgments.

3.6 Conclusion

Most diversity programs focus on historical prejudices (Von Bergen, Soper, & Foster, 2002) and assume that bias is intentional (Jackson & Joshi, 2010; Kalev, Dobbins, & Kelly, 2006). Future diversity training programs should use evidence from social and behavioral science on what bias is and how it functions. Moreover, science should be incorporated into the *evaluation* of diversity programs – as was in the current work – to clarify what the training does and does not do so that it can be adapted and improved to meet its goals.

Results from the current research showed that, by integrating scientific evidence in lieu of “guessing” what material might work best and delivering this material experientially rather than descriptively, automatic bias education is effective at changing individuals' *beliefs about bias* immediately after receiving it. Moreover, for some of the

beliefs, change continued to persist even months after. Thus, the original intervention was successful at making people aware of their biases (step 1 in Wilson and Brekke's contamination model). Nevertheless, despite making individuals' more aware of their biases, results showed that the automatic bias education was ineffective at changing participants' judgments and behaviors in which bias could be expressed (step 4 of the model). This may be due, in part, to the fact that participants lacked the motivation to assess the magnitude and direction of their biases (step 2 and 3 of the contamination model).

One possible way to help motivate people with the goal of reducing their biased responses is to incorporate experiential content focusing on how individuals can counteract the influence of automatic bias. Previous research has demonstrated that individuals are better at learning new material when it is given experientially rather than simply described (de Haan & de Ridder, 2006; DeRue & Wellman, 2009; Jarvis, 1987; Kolb, 1984; Pratt, 1993). Moreover, research has shown that individuals are more likely to be goal motivated when they believe the effort will be successful (Weiner, 1985). Thus, giving participants more experiential, rather than descriptive, knowledge on how to combat automatic bias might help individuals learn how reduce the effects of bias in future situations while simultaneously motivating them with the goal of doing so. For example, research has shown that when individuals are explicitly instructed to cognitively elaborate (e.g., thinking about counter-stereotypes) they show less automatic bias toward out-groups (for a review see Dasputa, 2009).

In addition, it is possible participants would benefit from follow-up sessions to help them pinpoint when, and in what situations, they are likely to express bias. For example, Rudman et al. (2001) showed that a semester long diversity course that provided repeated discussions and exercises on prejudice and conflict resulted in less anti-Black biases from its participants compared to a control group.

Finally, it is possible that integrating automatic bias with other effective forms of diversity training can help reduce bias. For example, the ultimate goal of multicultural education is to reduce ethnic, racial, and gender disparities. Using discussions, experiential exercises (e.g., role-playing), and intergroup contact, this type of program has shown some evidence in changing its participants attitudes and behaviors (for a review see Stephan, Renfro, & Stephan, 2004). As the present intervention showed evidence of changing awareness and motivation, combining its key features with multicultural education that may change attitudes and behaviors may be particularly potent. Thus, automatic bias education coupled with intergroup contact may address all components of the barriers to addressing bias in the Wilson and Brekke mental contamination model (1994).

By integrating these changes and incorporating experimental evaluation of the outcome, future interventions will be more successful in changing people's biased responses in decision making situations, ultimately meeting its goal of reducing bias.

References

- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal?: A field experiment on labor market discrimination. *American Economic Review*, 94, 991–1013.
- Carpusor, A. and Loges, W.E. (2006). Rental discrimination and ethnicity in names. *Journal of Applied Social Psychology*, 36, 934-952.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.
- Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163-70.
- Dalbert, C., Lipkus, I. M., Sallay, H., & Goch, I. (2001). A just and an unjust world: Structure and validity of different world beliefs. *Personality and Individual Differences*, 30, 561-577.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383.
- Dasgupta, N. (2009). Mechanisms underlying malleability of implicit prejudice and stereotypes: The role of automaticity versus cognitive control. In T. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination*. Mahwah, NJ: Erlbaum.
- de Haan, E., & de Ridder I. (2006). Action learning in practice: How do participants learn? *Consulting Psychology Journal: Practice and research*, 58, 216-231

- DeRue, D. S., & Wellman, N. (2009). Developing leaders through experience: The role of developmental challenge, learning orientation, and feedback availability. *Journal of Applied Psychology, 94*, 859-875.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Doise, W., & Mugny, G. (1984). The social development of the intellect. New York: Pergamon Press.
- Epley, N., & Dunning, D. (2000). Feeling “holier than thou”: Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*, 861–875.
- Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49*, 709-724.
- Flanagan C. A., & Faison, N. (2001). Youth civic development: Implications of research for social policy and programs. *Social Policy Report: Giving Child and Youth Development Knowledge Away, XV*, 3-14.
- 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.
- 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*, 1462-1480.

- Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- 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.
- Heesacker, M., Petty, R. E., Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality*, 51, 653-666.
- Hill, T. & Lewicki, P. (2006). Statistics methods and applications: A comprehensive reference for science, industry, and data mining. Tulsa, OK: Statsoft, Inc.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15, 635-650.
- Jackson, S. E., & Joshi, A. (2010). Work team diversity. In S. Zedeck (ed.), *APA Handbook of Industrial and Organizational Psychology, Volume 1*, pp. 651-686.
- Jarvis, P. (1987). *Adult Learning in the Social Context*. London: Croom Helm.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kassin, S. M., Sommers, S. R. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin*, 23, 1046-1054.

- Kalev, A., Dobbin, F., Kelly, E. (2006). Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies. *American Sociological Association*, 71, 589-617.
- Kerr, M., Sattin, H., & Trost, K. (1999). To know you is to trust you: Parents' trust is rooted in child disclosure of information. *Journal of Adolescence*, 22, 737-752.
- Kidder, D. L., Lankau, M. J., Chrobot-Mason, D., Mollica, K. A., & Friedman, R. A. (2004). Backlash toward diversity initiatives: Examining the impact of diversity program justification, personal and group outcomes. *International Journal of Conflict Management* 15, 77-104.
- Kolb, D. A. (1984). *Experiential Learning*, Englewood Cliffs, NJ: Prentice Hall.
- Kruglanski, A. W., & Mayseless, O. (1990). Classic and current social comparison research: Expanding the perspective. *Psychological Bulletin*, 108(2), 195-208.
- Levine, J. M., Resnick, L. B., & Higgins, T. E. (1993). Social foundations of cognition. *Annual Review of Psychology*, 44, 585-612.
- Lindner, N. M., Nosek, B. A., & Graser, A. (2009, February). *Age-based hiring discrimination as a function of equity norms and self-perceived objectivity*. Poster presented at the annual meeting of the Society of Personality and Social Psychology, Tampa, FL.
- Malos, S. (2007). Appearance-based sex discrimination and stereotyping in the workforce: Whose conduct should we regulate? *Employee Responsibilities and Rights Journal*, 19, 95-111.

- Menatti, A., Smyth, F. L., Teachman, B. A., & Nosek, B. A. (2011). Reducing stigma toward individuals with mental illnesses: A brief, online manipulation. Unpublished manuscript.
- McGurk, H & MacDonald, J. (1976). Hearing lips and seeing voices, *Nature*, 264, 746–748.
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice reduction efforts. *Journal of Personality and Social Psychology*, 65, 469-485.
- Monteith, M., J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases, *Social Cognition*, 19, 395-417.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, 83, 1029-1050.
- Naff, K. C., & Kellough, J. (2003). Ensuring employment equity: Are federal diversity programs making a difference? *International Journal of Public Administration*, 26, 1307-1336.
- Nagel, A. N., Joy-Gaba, J. A., & Nosek, B. A. (2010, April). Feeling objective enough to be biased: The effect of objectivity priming on gender bias in hiring decisions. Poster presented at the L. Starling Reid Undergraduate Psychology Conference, Charlottesville, VA.

- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565-584.
- Nosek, B.A. (2007). Understanding the individual implicitly and explicitly. *International Journal of Psychology*, 42, 184-188.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166-180.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265-292). New York: Psychology Press.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R., (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18, 36-88.
- Park, J. & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology*, 78, 1005-1023.

Pascarella, E. T., Edison, M., Amaury, N., Hagedorn, S., & Terenzini, P. T. (1996).

Influences on students' openness to diversity and challenge in the first year of college. *The Journal of Higher Education*, 67, 174-195.

Pickel, K. L. (1995). Inducing jurors to disregard inadmissible evidence: A legal explanation does not help. *Law and Human Behavior*, 19, 407-424.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832.

Pratt, D. D. (1993). Andragogy after twenty-five years. *New Directions for Adult and Continuing Education*, 57, 15-23.

Pronin, E., & Kugler, B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43, 565-578.

Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology*, 81, 639-656.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28, 369-381.

Ratcliff, J. J., Lassiter, G. D., Markman, K. D., & Synder, C. J. (2006). Gender differences in attitudes toward gay men and lesbians: The role of motivation to respond without prejudice. *Personality and Social Psychology Bulletin*, 33, 1325-1338.

- Schneider, W., & Schiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Shepard, R.N. (1990). Mind sights: Original visual illusions, ambiguities, and other anomalies. New York: WH Freeman and Company.
- Sidanius, J., & Pratto, F. (1999). Social dominance: An intergroup theory of social hierarchy and oppression. New York: Cambridge University Press.
- Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Snyder, M. (1987). *Public appearances, private realities: The psychology of self-monitoring*. New York: W. H. Freedman.
- Society for Human Resource Management (2010). Workplace diversity practices: How has diversity and inclusion changed over time? Retrieved October 12, 2010, from <http://www.shrm.org/Research/SurveyFindings/Articles/Pages/WorkplaceDiversityPractices.aspx>
- Stephan, C. W., Renfro, L., & Stephan, W. G. (2004). The evaluation of multicultural education programs: Techniques and a meta-analysis. In W. G. Stephan & W. P. Vogt (Eds.), *Education programs for improving intergroup relations* (pp.227-242). New York: Teachers College Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220-247.

- Sue, S., Smith, R. E., & Caldwell, C. (1973). Effect of inadmissible evidence on decisions of simulated jurors: A moral dilemma. *Journal of Applied Social Psychology, 3*, 345-353.
- Uhlmann, E.L., & Cohen, G.L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes, 104*, 207-223.
- Von Bergen, C.W., Soper, B., & Foster, T. (2002). Unintended negative effects of diversity management. *Public Personnel Management, 31*, 239-251.
- Walsh, W., Banaji, M. R., & Greenwald, A. G. (1995, June-July). *A failure to eliminate race bias in judgments of criminals*. Paper presented at the 7th Annual Convention of the American Psychological Society, New York.
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117-142.
- Weiner, B. (1985). An attributional theory of achievement. *Psychological Review, 92*, 548-573.
- Wheless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*, 250-257.

Appendix A

Self-Perception of Objectivity

My decision making is rational and objective.

When I make decisions, I am aware of the reasons for those decisions.

If my decision was influenced by unwanted biases, I would know about it.

Motivation to Address Automatic Bias

I am comfortable talking about potential biases I may have with other people.

The best method for preventing unwanted biases from influencing judgment of others is to have honest intentions about treating people fairly.

Awareness of one's own unwanted biases is an important step for reducing their influence on behavior.

I am motivated to ensure that unwanted biases do not influence my decision making.

Acknowledgement That Automatic Bias Exists

I am concerned that I may have unwanted biases that I am not aware of.

Everyone, including me, has biases toward other people.

Whether I like it or not, everyday experiences can influence my perceptions and assumptions about other people.

I am biased.

Whether I want to or not, I have more positive feelings about some racial groups than others.

Recognition that Automatic Bias can Influence Behavior

A significant portion of behavior is influenced by thoughts and feelings that people are not even aware of.

Whether I am aware of it or not, I use a person's race or ethnicity to form an impression of the kind of person they are.

Unwanted biases probably influence my decision making about other people.

If people try hard enough, their decisions about other people will not be influenced by unwanted biases.

Unwanted biases may affect the way I make decisions without my realizing it.

A "good person" is someone who does not have biases.¹¹

---

¹¹ Initially, I believed this item is an independent factor; thus, I did not anticipate that this item would load onto any of the hypothesized factors.

Appendix B

Bias Education Reminder Questionnaire

Control Conditions

Lady Gaga's Monster Ball Concert at JPJ on September 8, 2010

The University of Virginia versus Miami football game on October 30, 2010

Trick-or-treating on the Lawn on October 31, 2010

Convocation at JPJ on November 5, 2010

Reminder Condition

Lady Gaga's Monster Ball Concert at JPJ on September 8, 2010

The University of Virginia versus Miami football game on October 30, 2010

Trick-or-treating on the Lawn on October 31, 2010

The “Mindbugs” Lecture given by Dr. Brian Nosek at JPJ on August 23, 2010

Appendix C

Black Sounding Names Used in the Criminal Task

Darien Freeman
 Darnel Powell
 Luther Lyons
 Reggie Scott
 Jeroy Patterson
 Theo Thomas
 Leon Brown
 Lamar Rochester
 Terrell Tucker

Leotis Payne
 Ferrante Towns
 Ephraim Payton
 Everol Tubbs
 Percell Pinnock
 Leroy Washington
 Lamont Smith
 Kenyon Burrough
 Dwayne Townsend

White Sounding Names Used in the Criminal Task

Matthew Adams
 Mike Beatty
 Jack Connell
 Richard Glickman
 Eric Griffin
 David June
 Frank Logan
 Sam Meyers
 John Olson
 Allen Richards
 Daniel Stuart
 Timothy Wolf
 Bill Allking
 Scott Calvin
 Robert Forbes
 Abraham Gregory
 Bob Holmes
 Ed Lindsay
 Geoff McDonald
 Harry Newton
 Steve Perlin
 Brian St. John
 Peter Whittaker
 George Alcorn
 Joseph Brush
 William Deitch
 Greg Goldwater
 Michael Halloway
 Thomas Landorf
 Rob Ludgood
 Paul O'Gerow
 James Pierce
 Mark Wolenc

Sean Aleen
 Ralph Burke
 Tom Drake
 Steven Gollinger
 Teddy Hammick
 Stephen Lawson
 Jay McBride
 Fred Murray
 Jeff Pennebaker
 Donald Shirley
 Andrew Welz
 Craig Akers
 Henry Bostick
 Dennis Curban
 Kevin Goldin
 Dave Halbrook
 Roger Kraut
 Curt Lucas
 Ted Miller
 Arthur Palmer
 Ryan Rogers
 Jason Underwood
 Gary Woodson
 Joe Applegate
 Jim Carnes
 Victor Gaston
 Keith Griesing
 Leonard Jacobs
 Rich Livingston
 Andy Millton
 Art Payne
 Harvey Stone
 Johnnie Merritt

Appendix D

Belief in a Just World Scale

Justice prevails over injustice.

Injustices in all areas of life (e.g., professional, family, politics) are the exception rather than the rule.

People try to be fair when making important decisions.

People get what they deserve.

Basically, the world is a fair place.

In the long run, people will be compensated for injustice.

Appendix E

External Motivations to Reduce Prejudice

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

I try to hide any negative thoughts about other people in order to avoid negative reactions from my peers.

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

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

I try to act nonprejudiced because of pressure from others.

Internal Motivation to Reduce Prejudice Items

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

According to my personal values, using stereotypes about other social groups is OK.

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

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

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

Appendix F

Self Monitoring Scale

I try hard to imitate the behavior of others.

I'm not always the person I appear to be.

I would not change my opinions (or the way I do things) in order to please someone or win their favor.

I have considered being an entertainer.

I have never been good at games like charades or improvisational acting.

I have trouble changing my behavior to suit different people and situations.

At a party I let others keep the jokes and stories going.

I feel awkward in company and do not show up as well as I should.

I can look anyone in the eye and tell a lie with a straight face (if for the right end).

I may deceive people by being friendly when I really dislike them.

At parties and social gatherings, I do not attempt to do or say things that others will like.

I can only argue for ideas I already believe.

I can make impromptu speeches on topics I have almost no information.

I guess I put on a show to impress others.

I would probably make a good actor.

In a group I am rarely the center of attention.

In different situations and with different people, I often act like a different person.

I am not particularly good at making other people like me.

Appendix G

Unaware of Our Unawareness

Gavin Mandel^{12*},

“I’ll know it when I see it,” runs the popular refrain. It’s been used to explain how we can recognize everything from obscenity to true love. But how much can we trust what we see or, rather, what we think we see? For decades, cognitive psychologists have been discovering that there is more going on in our brains than we could ever be consciously aware of, even for a moment. The simple tasks of everyday life are so complex that they would overwhelm us if we had to supervise them all the time.

Consciousness Overthrown

What sorts of things influence our brains without our noticing? The list is extensive. The earliest of the true classics was a bystander intervention study (1). Imagine that you have just arrived at a psychology study. You are greeted by the experimenter and told that you will be having a group discussion via intercom. You’ll begin by taking turns talking about your experiences your freshman year, and then you’ll have a discussion. You’re told that you’ll be with just one other person. You’re shown into a room, the experimenter leaves, and the discussion begins. As the other person is talking, he begins to have a seizure, bumping around, speaking incoherently, and then falling ominously silent. What do you do?

Odds are, you think you’d ask for help. And you’re probably right. Now imagine the exact same scenario with just one minor change. Instead of being told that your discussion will involve you and one other person, you’re told that it will involve a group of four people. When the discussion begins, they introduce themselves and then the unfortunate subject has his medical crisis. Remembering that you can’t talk via intercom, what do you do? This could be life and death.

The answer is that you wait longer to ask for help, if you do at all. In fact, the odds of any one person helping are inversely proportional to the number of others who are in a position to help. Though researchers have found this effect in many studies, their research subjects tend to deny that it happens (2-5). They insist that the number of people in the study had not affected them whatsoever. Their brains had fooled them completely.

Recent Research

The most recent developments have been in the area of social behavior and goal pursuit. What is most striking about these demonstrations of unconscious influences is that the effects are obtained by extremely subtle psychological manipulations. A person is non-consciously cued to think about a certain idea by, for example, showing them adjectives related to rudeness in a “language skills test.” This is referred to as *priming*. If you take this person who is having unconscious “rude” thoughts and give him or her a chance to behave in a rude manner – such as interrupting an experimenter who is conversing with another person – the subject will exhibit greater rudeness, by interrupting faster (6). The subject does not realize that they’re acting differently because of the words they saw. They instead come up with a story about how they’d been having a bad day, or that they were in a hurry. Yet, time and time again, the nonconscious cues have their effect (7-9). In fact, we get the same results even if subjects see the “rudeness” words flash on a screen subliminally – i.e., so quickly they don’t realize they saw anything (10).

If simple behaviors can be nonconsciously cued, can motivations and goals? Priming the concept of achievement causes people to work harder and score higher on a verbal task (11), while making them think of cooperation causes them to act more cooperatively in economic simulations and Prisoner’s Dilemma games (12). People are unaware that these goals have been activated, even as they act in ways to attain those very goals (13-14). The nonconscious cue can even be as simple as sitting behind a professor’s desk, as opposed to in front of it (15). That can cue power and status (think about that in your next meeting with the boss). And all of these things happen without it ever occurring to us that we are acting any differently.

These studies are interesting, but maybe we are only so easily duped if we’re in experiments. But there have been other studies that have examined people who think the experiment is long over. In one study, subjects were led to think about elderly people. The subjects were then told that the experiment was over and that they were free to go. What did

---

¹² Department of Psychology

*To whom correspondence should be addressed.

Email: GMandel@Harvard.edu

Department of Psychology, Harvard University, Cambridge, MA 02138

the experimenter measure? How quickly the person walked down the hall when leaving the experiment. And yes, the people with the elderly thoughts took longer (16).

Real World Effects

There have been studies on the impact of aggressive cues like weapons on later aggressive behavior (17), the impact of exposure to television violence on aggressiveness in children (18-23), and the impact of witnessing helpful acts on later probability of helping a person in need (24-25). Those studies showed consistently strong influence by context on behavior in real world settings. Rarely do any of the people in those studies think they were influenced. After all, how many teenagers think that playing Grand Theft Auto will make them more violent? Yet research suggests it does (26).

Modern American popular culture and media have created a strong association between African American people and bad acts (27). And some well meaning people don't react well when they encounter African Americans. It is nothing intentional, they may just not quite meet the black person's eyes, or might not say hello, or might just have a feeling of discomfort (28). It is the little things, except the little things add up to Big Things, like being less likely to offer black candidates jobs in interview simulations, or ending a simulated loan interview more quickly (29). This is called Aversive Racism. Unlike its more overt ancestors, this form is almost entirely unconscious. In fact the person is often horrified if shown evidence that they reacted differently to black and white people (30). Aversive Racism can be measured by a host of measures, the most famous of which is the Implicit Attitudes Test. That reaction time task shows how closely concepts are linked in the brain. Several years ago, researchers at the Cognitive Science Laboratory at MIT began to trace brain activations in people who rated high in Aversive Racism. Different areas of their brain "light up" when looking at a black face than a white one (31). Some of those areas are linked to basic fear responses, and some involve more complex cognitive activation (32). In either case, people are unaware that they respond differently to the different faces. They often hold political and social beliefs that are incredibly accepting and tolerant. Yet still, the data show a reliable and unconscious prejudice. Often we, as scientists, have trouble accepting that we too could be biased, but we have no choice but to admit the possibility.

Aversive Racism is but one example of our tendency to make snap judgments about other people without realizing it. Another classic example is the *halo effect*, whereby our first impressions are unintentionally influenced by a single positive trait. For example, physically attractive people are also judged to be more likeable, social, and talented (33). Because this effect happens without our intending it (or even knowing we did it), we tend to be overconfident in the accuracy of first impressions.

Lifting the Hood

So how does our conscious mind deal with all of these non-consciously motivated behaviors? As suggested by the rudeness study mentioned earlier, it tries to interpret them, to tell a story around them. The classic pantyhose study is a good example. People were brought into a lab and asked to choose between several different sets of pantyhose. The participant would always claim that the pair they chose was somehow "better." They said it was silkier, they said it was sheerer, they said a great many things. How do we know that they were mistaken? Well, each of the pairs was exactly the same, but they always picked the pair furthest to the right, even though the pairs were shuffled every time (33). They picked that pair because they saw it first as they entered the room, but they didn't know that was the reason. Their minds were left to struggle for some other grounds on which to retroactively base their decision. We do think this way everyday. Ever stopped to wonder why you did something? Its hard to know if you found the "real" reason.

The key is efficiency. The title of one recent book puts it clearly, we are gifted and cursed with "The Cognitive Shortcuts That Make Us Smart" (34). Our brains have found clever ways of preserving processing power while still getting it right enough of the time. What is enough? Well, more now than it used to be. The brain wasn't designed with Mass Media in mind. As the last two decades have seen the rise of evolutionary theories of psychology, it is natural to think of these effects in terms of what is adaptive. Our brains do a lot of things without getting our full informed consent, but they only have our best interests in mind.

Unaware of Our Unawareness

If it isn't enough that our consciousness can be so easily undermined, we are also left unaware of its failure. Try as they might, experimenters have not been able to make people realize when they have been unintentionally influenced. We believe that we would know if our actions were being so drastically altered. Yet these studies and hundreds of others have all shown that we know less about our motivations and about the sources of our actions, judgments, and decisions than we thought. Indeed, our intuitions may provide little but false confidence.

References and Notes

1. Darley, & Latane (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
2. Horowitz, (1971). The effect of group norms on bystander intervention. *Journal of Social Psychology*, 83, 265-273.
3. Christy (1994). Bystander responses to public episodes of child abuse. *Journal of Applied Social Psychology*, 24, 824-847.
4. Tice, & Baumeister (1985). Predicting the bystander effect. *Journal of Personality and Social Psychology*, 49, 420-428.
5. Bryan, & Test (1967). Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology*, 6, 400-407.
6. Bargh, Chen, & Burrows (1996). Automaticity of social behavior: effects of priming on action. *Journal of Social Psychology*, 81, 230-244.
7. Johnson, & Tanya (1999). The cognitive monster: The case against controllability of automatic effects. *Theories in Social Psychology*, 361-382.
8. Grubman, & Artez (1999). The most powerful manipulative messages are hiding in plain sight. *Applied Psychology*, 57, 6-19.
9. Milstein (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462-479.
10. Bargh et al (1996). Experiment 3.
11. Gollwitzer, Barndollar, Troetchel, (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, 81, 1014-1027.
12. Neuberg (1988). Behavioral implications of information presented outside of conscious awareness: The effect of subliminal presentation of trait information on behavior in the prisoner's dilemma game. *Social Cognition*, 6, 207-230.
13. Fitzsimons & Grand (2002). Nonconscious motivations: their activation, operation, and consequences. *Self and Motivation*, 28, 13-41.
14. Nisbett & Wilson (1977a). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231-259.
15. Chen, Lee-Chai (2001). Relationship orientation and the effects of social Power. *Journal of Personality and Social Psychology*, 80, 173-187.
16. Bargh et al. 1996, experiment 2.
17. Berkowitz, & Leipage, (1967). Weapons as aggression-eliciting stimuli. *Journal of Personality and Social Psychology*, 7, 202-207.
18. Donnerstain, Slaby, & Eron, (1994). The mass media and youth aggression. In Eron, Gentry, & Schlegel, *Reason to Hope: A psychosocial perspective on violence and youth*. Washington DC, American Psychological Association.
19. Anderson & Bushman (2002). The effects of media violence on society. *Science*, 295, 2377-2379.
20. Paik & Comstock (1994). The effects of television violence on antisocial behavior: a meta-analysis. *Communication Research*, 21, 516-546.
21. Wood, Wong, & Chachere (1991). The effect of media violence on viewer's aggression in unconstrained social Interaction. *Psychological Bulletin*, 109, 371-383.
22. Meyer (1972). The effects of justified and unjustified real film violence on aggressive behavior. *Journal of Personality and Social Psychology*, 23, 21-29
23. Johnson, Cohen, Smailes, Kasen, & Brook (2002). Television viewing and aggressive behavior during adolescence and adulthood. *Science*, 295, 2468-2471.
24. Bryan & Test (1967). Models and helping: naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology*, 6, 400-407.
25. Piliavan & Callero (1991). *Giving blood: the development of an altruistic identity*. Baltimore: Johns Hopkins University Press.
26. Johnson & Bushman (2004). Aggression modeling and popular video games. *Journal of Personality and Social Psychology*, 90, 186-205.
27. Monteith, Sherman, & Devine (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review*, 1, 62-82.
28. Nail, Harton, & Decker (2003). Political orientation and modern versus aversive racism: tests of Dovodio and Gaertner's (1988) Integrated Model. *Journal of Personality and Social Psychology*, 84, 754-770.
29. Word, Zanna, & Cooper, (1974). The nonverbal mediating of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10, 109-120.
30. Dovodio & Gaertner (2004). Aversive Racism. In Zanna (ed) *Advances in Experimental Social Psychology*. New York: Academic Press.
31. Cohen, Goren, Harris & Fiske (2004). Brain activation and stereotype salience: neurological implications of unconscious racism. *Journal of Cognitive Science*, 9, 300-314.
32. Cohen et al, 2004,
33. Nisbett & Wilson (1977b). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250-256.
34. Nisbett & Wilson (1977a).
35. Gigerenzer, Todd, and ABC Research Group (1999). *The cognitive shortcuts that make us smart: evolution and cognition series*. Oxford University Press.

Appendix H

Bias Blindspot Questionnaire

Psychologists have claimed that some people show a “disconfirmation” tendency in the way they evaluate research about potentially dangerous habits. That is, they are more critical and skeptical in evaluating evidence of danger when they personally engage in such behavior than when they do not personally engage in such behavior.

Psychologists have claimed that some people show a “rationalizing” tendency in the way they feel about things they buy. That is, after choosing to buy a particular item, and thereby foregoing other alternatives, people are inclined to see the item they bought as higher in quality, more attractive in its features, and a better “deal” than the alternatives.

Psychologists have claimed that some people show a “self-interest” effect in the way they view candidates in elections. That is, people’s assessments of qualifications, and their judgments about the extent to which particular candidates would pursue issues and policies good for the voters as a whole, are influenced by their feelings about whether the candidate would serve their own particular political, economic and social interests.

Psychologists have claimed that some people show a “self-serving” tendency in the way they view their academic or job performance. That is, they tend to take credit for success but deny responsibility for failure; they see their successes as the result of personal qualities, like drive or ability, but their failures as the result of external factors, like unreasonable work requirements or inadequate instruction.

Psychologists have claimed that some people show a tendency to make “overly dispositional inferences” in the way they view victims of assault crimes. That is, they are overly inclined to view the victim’s plight as one he or she brought on by carelessness, foolishness, misbehavior, or naiveté.

Psychologists have claimed that some people show a “halo” effect in the way they form impressions of attractive people. That is, when it comes to assessing how nice, interesting, social, able, etc. someone is, people tend to judge an attractive person more positively than he or she deserves.

Psychologists have claimed that some people show a “hostile media” effect in the way they view television or newspaper coverage of controversial or partisan issues. That is, they tend to perceive neutral media reports as presenting an inaccurately favorable view of the other side, and an inaccurately negative view of their own side. As a result, they see the media as “hostile,” or “biased against” their side.

Psychologists have claimed that some people show an “anchoring” effect in the way they form impressions of people. That is, people tend to form strong first impressions of individuals they meet, and then they tend to make only minor adjustments from those initial impressions.

Psychologists have claimed that some people show a “better-than-average” effect in the way they judge their own traits and abilities. That is, when it comes to positive traits, people tend to rate themselves as having these more than average, and when it comes to negative traits, they tend to rate themselves as having these less than average.

Psychologists have claimed that some people show an “ingroup-outgroup” tendency in the way they treat people who are members of different groups. That is, people tend to judge people who share their own group identities (such as people from their own hometown, ethnic background, or club) more positively than they judge people from other groups.

Supplement

Study 4 Additional Analyses

I Have Bias. An ANOVA revealed no significant differences based on condition, $F(3, 445) = 0.35, p = 0.79, d = 0.06$. In the follow-up contrast analyses, the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 447) = 0.43, p = 0.51, d = 0.06$, nor did the bias interventions show significantly differ compared to the control condition, $F(1, 447) = 0.24, p = 0.62, d = 0.05$.

Automatic Bias Influences Me. Results showed no significant differences based on condition, $F(3, 445) = 0.80, p = 0.50, d = 0.08$. Planned contrasts revealed that the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 447) = 0.14, p = 0.71, d = 0.04$, nor did the bias interventions show significantly differ compared to the control condition, $F(1, 447) = 0.96, p = 0.33, d = 0.09$.

Self-Perception of Objectivity. Results revealed a significant effect of condition, $F(3, 445) = 3.03, p = 0.03, d = 0.17$, although post-hoc tests did not reveal any significant differences by condition. Contrast analyses revealed that the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 447) = 2.83, p = 0.09, d = 0.16$, nor did the bias interventions show significantly differ compared to the control condition, $F(1, 447) = 2.23, p = 0.14, d = 0.14$.

Motivation to Address Automatic Bias. AN ANOVA revealed no significant differences based on condition, $F(3, 445) = 1.39, p = 0.25, d = 0.11$. Contrast analyses revealed that the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 447) = 0.30, p = 0.58, d = 0.05$, nor did the bias interventions show significantly differ compared to the control condition, $F(1, 447) = 0.05, p = 0.82, d = 0.02$.

Knowing and Preventing Automatic Bias. Results showed no significant difference based on condition, $F(3, 445) = 2.08, p = 0.10, d = 0.14$. Planned comparisons revealed that the bias education treatment condition was not different from the other conditions (i.e., IAT, Bias Blindspot, and control), $F(1, 447) = 0.78, p = 0.38, d = 0.08$, nor did the bias interventions show significantly differ compared to the control condition, $F(1, 447) = 2.92, p = 0.09, d = 0.16$.

Other Dependent Variables

Behavioral Intention and Policy Support Questionnaire. An ANOVA revealed a no significant effect of condition on participants responses, $F(3, 445) = 1.71, p = 0.17, d = 0.12$.

Planned contrasts revealed the bias education did not significantly change participants' intentions or support compared to the other conditions, $F(1, 447) = 2.69, p = 0.10, d = 0.16$. Similarly, contrast analyses showed that there was no significant differences between participants receiving a bias intervention and control participants, $F(1, 447) = 0.39, p = 0.53, d = 0.06$.

Bias Blindspot Questionnaire. An ANOVA revealed a significant effect of condition on participants responses to the blindspot questionnaire, $F(3, 447) = 2.62$, $p = 0.05$, $d = 0.15$. Post-hoc tests revealed that participants who received the shortened automatic bias education reported higher agreement that they were susceptible to automatic processes ($M = 3.76$, $SD = 0.72$) than participants who had received the bias blindspot article ($M = 3.50$, $SD = 0.73$).

Planned contrasts revealed the participants who received the bias education reported significantly higher agreement that they were susceptible to automatic processes compared to the other conditions, $F(1, 449) = 7.51$, $p = 0.01$, $d = 0.26$. Contrast analyses also showed that there was no significant differences between participants receiving a bias intervention and control participants, $F(1, 449) = 0.68$, $p = 0.41$, $d = 0.08$.

Criminal Task. A 2 (Target Race: Black versus White) x 4 (Condition: Bias Education versus IAT versus Bias Blindspot versus Control) ANOVA was conducted on the proportions of names identified as criminals. Across conditions, Black sounding names were more likely than White sounding names to be misjudged as criminals, $F(1, 443) = 9.58$, $p = 0.002$, $d = 0.11$. Results revealed no significant interaction of Target Race X Condition for participants' criminal judgments of Black and White sounding names, $F(3, 443) = 1.63$, $p = 0.18$, $d = 0.06$ indicating that the tendency to misjudge Black sounding names as criminal did not decrease as a function of receiving bias education.

Study 5 Additional Analyses

Automatic Racial Attitudes. Overall, participants automatically preferred White people to Black people ($M = 0.36$, $SD = 0.42$), $t(363) = 16.40$, $p < 0.0001$, $d = 0.86$. A t -test revealed that the shortened bias education did not influence participants' automatic attitudes; educated participants similar pro-White bias ($M = 0.37$, $SD = 0.44$) to control participants ($M = 0.35$, $SD = 0.41$), $t(362) = 0.45$, $p = 0.65$, $d = 0.05$.

Self-Reported Racial Attitudes. Overall, participants explicitly preferred White people to Black people ($M = 0.60$, $SD = 1.01$), $t(367) = 11.49$, $p < 0.0001$, $d = 0.60$. Paired t -tests revealed that participants reported significantly higher feelings of warmth toward White people ($M = 6.85$, $SD = 1.57$) than Black people ($M = 6.38$, $SD = 1.69$), $t(368) = 5.84$, $p < 0.0001$, $d = 0.30$. A t -test revealed that bias educated participants reported similar preferences for White people to Black people ($M = 0.63$, $SD = 1.12$) compared to control participants ($M = 0.58$, $SD = 0.92$), $t(366) = 0.47$, $p = 0.64$, $d = 0.05$.

Study 6 Additional Analyses

Automatic Racial Attitudes. Overall, participants automatically preferred White people to Black people ($M = 0.37$, $SD = 0.43$), $t(734) = 23.31$, $p < 0.0001$, $d = 0.86$. An ANOVA revealed that participants who received the shortened bias education showed similar pro-White bias ($M = 0.40$, $SD = 0.42$) to participants who viewed the dancing video ($M = 0.34$, $SD = 0.44$) and "The Last Lecture" video ($M = 0.35$, $SD = 0.44$), $F(2, 732) = 1.32$, $p = 0.27$, $d = 0.08$.

Self-Reported Racial Attitudes. Overall, participants explicitly preferred White people to Black people ($M = 0.50$, $SD = 1.03$), $t(745) = 13.29$, $p < 0.0001$, $d = 0.48$. Paired t -tests revealed that participants reported significantly higher feelings of warmth toward White people ($M = 6.92$, $SD = 1.58$) compared to Black people ($M = 6.44$, $SD =$

1.66), $t(748) = 8.18$, $p < 0.0001$, $d = 0.30$. An ANOVA revealed that participants who received the bias education reported similar explicit preferences ($M = 0.50$, $SD = 1.08$) compared to participants who viewed the dancing video ($M = 0.43$, $SD = 0.94$) and “The Last Lecture” video ($M = 0.59$, $SD = 1.03$), $F(2, 732) = 1.21$, $p = 0.30$, $d = 0.08$.