

**Reducing Discrimination:  
A Comparative Investigation of Six Intervention Strategies**

Geneva Juanyu Yang

PSYC380 Honours Research Project Seminar

Supervisor: Dr. Jordan Axt

Department of Psychology

McGill University

March 14th, 2021

### **Abstract**

Discrimination occurs when people intentionally or unintentionally incorporate irrelevant social information into decision making. Different interventions have been developed to reduce such biases in social judgment. However, past work has lacked standardization across samples and outcomes, making conclusions about the comparative effectiveness of these interventions difficult. In two pre-registered studies (total  $N > 6,000$ ) that used the same outcome measure, we drew from prior literature to develop and compare six possible discrimination-reducing interventions: imposing accountability, reviewing a lesson on confirmation bias, requiring a delay in response, creating implementation intentions, committing to objectivity, and adding financial incentives. Participants evaluated applicants to a hypothetical honor society using relevant academic credentials, while also viewing faces known to elicit favoritism based on physical attractiveness. In both studies, the implementation intentions intervention and the delay intervention significantly reduced discrimination, albeit through different mechanisms. This work provides the largest comparative assessment of strategies to reduce biased judgment and advances understanding of how and when various interventions may reduce discrimination.

*Keywords:* discrimination, decision-making, bias, social judgment, judgment bias task, signal detection

## Introduction

Discrimination can be defined as the differential treatment of people based on their membership in a social group (Axt & Lai, 2019). Compared to stereotypes, which are cognitive associations between groups and certain attributes, and prejudices, which are affective attitudes toward certain groups, discrimination is the behavioral output that generates and sustains unjustified group-based disparities in real-world outcomes (Fiske, 1998). Both field and laboratory research shows that discrimination is pervasive in the general population as well as among professionals. For example, a review of laboratory studies on racial discrimination (Crosby et al., 1980) revealed that (a) White people help other White people more than Black people; (b) under sanctioned conditions, White people display more aggressive verbal and non-verbal behaviors toward Black people; and (c) when interacting with Black people, White people exhibit non-verbal cues that indicate discomfort or hostility, but when asked to report on their attitudes, White people report no differential treatment. More recently, a great deal of research has revealed that discrimination can often occur subtly, manifesting in covert and possibly automatic ways, where a gap exists between people's self-reported egalitarian beliefs and their unrecognized discriminatory behaviors (Dovidio & Gaertner, 2000; Moss-Racusin et al., 2012).

Whether intended or unintended, discrimination on the basis of such ostensibly irrelevant characteristics like race, ethnicity, gender, sexual orientation, and physical appearance has a significant impact on real world outcomes, including hiring (Shaffer et al., 2000; Hosoda et al., 2003), firing (Commisso & Finkelstein, 2012), healthcare (Romanelli & Hudson, 2017), housing (Kattari et al., 2016; Ross & Turner, 2005), and lending (Ross & Yinger, 2002). For instance, in one field experiment conducted in India (Thorat & Attewell, 2007), researchers applied to over 600 job postings, each with three fictitious curriculum vitae that were equivalent on outcome-

relevant criteria — such as educational attainment — but differed in their names; across conditions, companies received a resume with a name suggesting that the applicant was either Dalit (a socially marginalized group that has the lowest ranking in the Hindu caste system), a Muslim (another marginalized group in Indian culture), or a Brahmin (a well-respected group that has the highest ranking in the Hindu caste system). Although applicants had the same qualifications, results found that the Dalit and the Muslim applicant received less interview offers than the Brahmin. Using a similar approach, a study in Sweden (Rooth, 2009) created two fictitious job applicants who differed only on physical appearance (headshots are often included in job applications in Europe), such that the photo of one applicant was the photoshopped version of another applicant, only made to look obese. Results found that the more obese applicant received significantly fewer interview offers compared to the non-obese applicant.

Given the real-world impact of such discrimination, developing, testing and comparing interventions that reduce discrimination is a crucial and pressing issue. While field and audit studies, like the two examples described above, are helpful in understanding the real-world implications of discrimination, it would be costly and difficult to test multiple interventions in a field setting. As a result, it may be more efficient and productive to use controlled, laboratory settings to refine interventions that can then be deployed in the field.

However, one clear limitation of laboratory studies on this topic is the difficulty in capturing discriminatory behavior. Existing measures of bias, such as the Implicit Association Test (Greenwald et al., 1998), can easily serve as the outcome of interest in lab-based interventions (e.g., Lai et al., 2014), but the relevance of measures like the IAT to discriminatory behavior is contested (Forscher et al., 2019). It is thus a major challenge to be able to develop a

lab-based measure that simulates a context where discriminatory behavior can be directly studied.

A promising development in this area comes from a recent measure, the Judgment Bias Task (JBT; Axt et al., 2018). In a JBT, people evaluate different profiles with relevant or irrelevant criteria for a particular outcome. In one version of the JBT, participants are on the selection committee for a hypothetical academic honor society and are given a series of applicant profiles to accept or reject. The profiles contain both relevant information, such as GPA and interview score, and irrelevant social information, such as political affiliation, race, and/or a photo signaling physical attractiveness. Bias is often measured using a within-subjects approach by adopting a Signal Detection Theory analysis in comparing the criterion values for accepting applicants from the two social groups represented in the task (e.g. Black people and White people); here, a lower criterion value for an acceptance decision towards one group over another group indicates that despite the two groups having the same overall level of qualifications, applicants from one group were held to a more lenient standard for admission. Furthermore, when coupled with self-report measures, researchers have found that those participants who after the task reported not wanting to use social information in their judgments, or reported *having* not used social information in their judgments, still showed biases in the task on average, indicating that the JBT is capable of detecting discrimination that exists outside of conscious awareness or control (Axt et al., 2018; Axt & Lai, 2019).

### **Prior Research on Interventions to Improve Decision Making**

The present work uses the JBT to investigate what interventions best reduce discrimination. Although multiple narrative reviews and meta-analyses have compared the effectiveness of various de-biasing strategies (Paluck & Green, 2009; Fitzgerald et al., 2019),

there is a lack of comparative and standardized experimental research on reducing socially biased decision-making using the same sample source and outcome measure. Having the same sample source and outcome measure may provide clearer evidence for the relative effectiveness of each intervention, as meta-analytic reviews must often collapse across studies that use varying populations and outcomes. Indeed, the heterogeneity of samples and outcomes used in such meta-analyses may make the conclusions less reliable when compared to a standardized experimental study (Lee, 2019). To gain traction on this issue, we identified six interventions that were used in prior psychological research that were found to either improve reasoning or decision-making, though not all have been applied to social judgment specifically.

### *Accountability*

Accountability refers to the expectation that one may need to justify their beliefs, attitudes, and actions to others (Lerner & Tetlock, 1999). There are different ways to impose accountability, including evaluation (participants expect that their performance will be assessed according to some rules), reason-giving (participants expect that they must provide reasons for their behaviors or attitudes), or the mere presence of another person. Multiple studies have found that imposing accountability significantly reduced judgment biases, but only if they were imposed before completing the task and if the biases were attributed to the failure to use all relevant cues and critically attend to one's decision processes.

The psychological mechanism behind these accountability interventions may be that impression management processes triggered by accountability affect cognitive processes, resulting in preemptive self-criticism that partly shields people from mindlessly applying simple heuristics (Lerner & Tetlock, 1999). For example, the recency effect (a cognitive bias where more recent information is given greater weight in judgment) was eliminated when

accountability was imposed on M.B.A students who were asked to judge the likelihood that a firm would fail (Kennedy, 1993). In the study, several statements about the firm were presented, with half of them supporting viability and half of them supporting failure. Participants in a control condition were more influenced by whatever information was presented last, whereas those undergoing an accountability manipulation (being told that their response would be evaluated by a group of experts and that they would need to provide reasons to justify their judgment) exhibited no such recency effect.

### ***Confirmation Bias***

Confirmation bias is the tendency to seek out or interpret information that is consistent with existing beliefs. For example, in the JBT, when participants already believe a student to be qualified, they might focus on the student's high interview score while ignoring their low GPA and low recommendation letter strength. Alternatively, if participants believe a student to be unqualified, they might focus on the student's low GPA while ignoring their high interview score and recommendation letter strength. Therefore, understanding and suppressing confirmation bias may be helpful in improving judgment.

In a recent study (Sellier et al., 2019), participants received an intervention bundle consisting of playing a detective video game designed to inhibit confirmation bias, reflecting on strategies to mitigate the bias, and completing practice problems that illustrated the bias. A complex business case that typically elicits confirmation bias was then presented to the students, who were asked to make a decision. Results found that the treatment group showed a 29% reduction in the likelihood of exhibiting a confirmation bias in their decision-making. These findings suggest that learning about confirmation bias before the evaluation process in the JBT may reduce biased judgment.

### ***Delay***

Rational, unbiased decision-making takes time, and allocating less time than is needed, or is perceived to be needed, causes “time stress” (Ariely & Zakay, 2001). The consequences of time stress are mostly negative, including increased tendency to use simple heuristics, reduced information search and processing, reduced range of alternatives and dimensions considered, forgetting of important information, and wrong judgment and evaluation (Ariely & Zakay, 2001). Under time stress, people resort to using different decision strategies that are likely to reduce accuracy and increase bias. In an experiment conducted by Axt and Lai (2019), greater time pressure when completing the First Person Shooter Task (FPST), a decision-making task where participants try to quickly identify guns or harmless objects in the hands of Black or White targets, caused participants to exhibit stronger racial bias in judgment. In another experiment, requiring a 4500 ms delay in making a decision on the JBT increased accuracy in identifying more qualified candidates compared to participants in an untimed control condition, and this increased overall accuracy translated into reduced discrimination (Axt & Lai, 2019).

### ***Implementation Intentions***

Implementation intentions are a self-regulatory strategy that uses “if-then” statements to help achieve certain goals. Rehearsing concrete distraction-inhibiting if-then statements to avoid biasing information has been shown to improve decision-making. In one study (Mendoza et al., 2010), participants were asked to complete an FPST, but before completing the task, half of the participants rehearsed a distraction-inhibiting implementation intention (“If I see a person, then I will ignore his race!”). The result showed that the implementation intentions group was more accurate and showed less racial bias than the control group.



A possible explanation for this phenomenon comes from the notion of reflexive action control (Amodio et al., 2007), where people automatically initiate goal-directed behaviors without conscious reflection. By linking a situational cue (if) with goal-directed behavior (then), the planned response can be implemented without deliberation, which may be particularly useful to combat judgment biases that are at least partly automatic. Given this prior work, rehearsing distraction-inhibiting implementation intentions may be effective in reducing bias in the JBT.

### ***Objectivity***

Asking participants to commit to being unbiased and listing desired criteria in advance may create a need for greater consistency between one's values and actions. Some evidence for the effectiveness of committing to objectivity can be found in Uhlmann and Cohen (2005). Here, participants were asked to evaluate a candidate for police chief that was either male or female, and the applicants' strengths were either described as being streetwise or well-educated. In control conditions, results found that the hiring criteria emphasized in judgment (either being streetwise or well-educated) were consistent with whatever traits were shown on the profile of the male applicant, meaning participants favored the male applicant over the female applicant regardless of actual qualifications. However, when participants were asked to rate the importance of each criteria *prior* to learning about the applicants' gender, there was no significant difference in the evaluation of male and female applicants.

Committing to criteria beforehand may then be helpful when judgments are ambiguous. Much like the outcome used in Uhlmann and Cohen (2005), the JBT also evaluates profiles containing ambiguous information about an applicant's strength, as there are four different relevant criteria to evaluate. By explicitly committing to using the objective criteria beforehand, participants may have less biased judgment on the JBT.

### ***Reward***

A final intervention strategy may be a simple reward for accurate, unbiased behavior. In a study conducted by Stone and Ziebart (1995), undergraduates who were offered performance-contingent financial incentives spent significantly more time on their decisions, examined more information, and made more accurate choices compared to those who were offered random financial rewards (i.e., not tied to performance). In another study (Tosi et al., 1997), participants were put in an experimental case study scenario, and they chose the profit-maximizing strategy more when their rewards were contingent on the outcomes of their performance. This prior work demonstrates that the opportunity to receive a performance-contingent financial incentive may increase motivation to regulate bias and improve performance.

### **The Present Work**

The primary aim of this work is to test and compare the effectiveness of the six identified interventions in reducing the magnitude of discrimination using the JBT. In addition, we seek to replicate previous studies that have shown the presence of physical attractiveness bias in admission and hiring (Hosoda et al., 2003; Axt et al., 2018; Axt et al., 2019). We will also conduct exploratory analysis on the effect of the six interventions on measures of explicit attitudes, implicit attitudes, desired performance, and perceived performance.

According to a series of studies conducted by Axt and Lai (2019), the magnitude of discrimination is shown to depend on two empirically independent measures: noise and bias (Green & Swets, 1966). Noise is the amount of errors made in judgment, resulting from an inability to integrate and evaluate multiple criteria. In the JBT, noise is measured using sensitivity ( $d'$ ), which is one's ability to identify the more over the less qualified applicants.

Increasing sensitivity then reduces discrimination by lessening the *total* number of errors made, but not necessarily impacting the proportion of remaining errors that favor a particular group.

Alternatively, bias is the degree to which errors disproportionately favor one group over another, which results from greater use of irrelevant social information in judgment. In the JBT, bias is measured by comparing the difference in criterion for accepting applicants from two different social groups. A lower criterion for one social group means that group received a greater proportion of “beneficial” errors (falsely admitting less qualified applicants), whereas the group with the higher criterion received comparatively more “detrimental” errors (falsely rejecting more qualified applicants). Reducing criterion bias decreases discrimination by lessening the relative rate at which errors favor one group over another, though this does not necessarily translate into fewer errors overall. Given these two distinct paths to reducing discrimination, we evaluated each intervention on its ability to increase sensitivity (i.e., reduce noise) and lessen relative gaps in criterion (i.e., reduce bias).

## **Method**

### **Open Science Practices**

We reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the pre-registration and study materials that are available at <https://osf.io/fq4vb/> for Study 1 and <https://osf.io/mrty4/> for Study 2.

### **Participants**

In Study 1, 4011 volunteer participants from the Project Implicit research pool (64.5% female, 68.7% White,  $M_{\text{Age}} = 35.25$ ,  $SD = 15.40$ ) completed at least the JBT. Participants were excluded from analysis if they accepted less than 20% of applicants or more than 80% of applicants, and if they accepted or rejected every more or less physically attractive applicant

(Axt, Nguyen & Nosek, 2018). In Study 1, this meant 288 participants were excluded, which left 3723 eligible participants in total (Control N = 545, Accountability N = 554, Confirmation N = 548, Delay N = 531, Implementation N = 507, Objectivity N = 500, Reward N = 538). This sample size provided greater than 90% power for detecting a between-subjects effect as small as Cohen's  $d = .225$ .

In Study 2, 2714 participants were recruited from Prolific and were rewarded £1 for the completion of the study (39.2% female, 83.8% White,  $M_{\text{Age}} = 26.90$ ,  $SD = 8.96$ ). Following the same exclusion criteria as in Study 1, 145 participant were excluded, which left 2569 eligible participants in total (Control N = 343, Accountability N = 395, Confirmation N = 367, Delay N = 402, Implementation N = 338, Objectivity N = 349, and Reward N = 375). This sample size provided greater than 90% power for detecting a between subjects effect as small as Cohen's  $d = .22$  (given the one-sample tests that were specified in our pre-registration).

## **Procedure**

Participants in Study 1 completed four study components in the following order: Participants first received the bias-reduction intervention (if there was one), then completed the JBT, followed by measures of perceived performance, desired performance, and explicit attitudes, and finally a Brief Implicit Association Test (BIAT) assessing implicit evaluations of more and less physically attractive people.

Participants in Study 2 completed the same components in the same order with the exception that they did not complete the BIAT.

## ***Experimental Conditions:***

Before completing the JBT, participants were randomly assigned to one of seven conditions. Please see Appendix 1 for the exact instructions given in each condition.

1. Control: Participants received no additional information beyond the typical JBT instructions.
2. Accountability: Participants were warned that their responses will be reviewed and analyzed by a panel of researchers that are experienced in evaluating students.
3. Confirmation Bias: Participants read a brief lesson about confirmation bias and how to combat the bias by adopting a 'hypothesis-disconfirming' strategy — looking for information that suggests a student may be unqualified, rather than only looking for information that suggests a student may be qualified.
4. Delay: Participants were told that they would complete, and then completed, a version of the JBT that has a four-second delay before responses are allowed.
5. Implementation Intentions: Participants were asked to rehearse, recite, and use the strategy "If I see a student's application, then I will ignore their face."
6. Commitment to Objectivity: Participants were informed that deciding on a set of parameters to use before beginning the task could prevent them from being distracted by irrelevant information during the task. They then wrote about how they would complete the task objectively, and why they believed they could effectively do so.
7. Reward: Participants selected a charity that would receive a \$5 donation if they were in the top 10% of accuracy (the payments were made at the study's completion).

### ***Academic Judgment Bias Task***

Participants were asked to make decisions on acceptance into a hypothetical academic honor society for 64 applicants. Each applicant profile contained four pieces of relevant information: Science GPA (on a scale of 1-4), Humanities GPA (on a scale of 1-4), letter of recommendation quality (with four categories: poor, fair, good, excellent), and interview score

(on a scale of 1-100). Each profile also contained a photo of the applicant designed to elicit a physical attractiveness bias. Participants were asked to accept approximately half of the applicants.

Profiles were constructed such that half of them were quantifiably more qualified than the other half. Qualification was calculated by converting each piece of information to a 1 to 4 scale. The GPAs were already out of 4. The four categories of the recommendation letter quality were converted to numbers (poor = 1, fair = 2, good = 3, excellent = 4). The interview scores were divided by 25. These four numbers were then summed up to yield a score of either 13 or 14. Profiles with a score of 14 were considered more qualified and those with a score of 13 were considered less qualified.

The photos of the profiles were selected such that there was an equal number of male and female faces, and all of them were White and smiling. These faces were pretested to vary on physical attractiveness ( $d = 2.64$  in attractiveness ratings of the more versus less physically attractive stimuli sets in Axt et al., 2018). Each participant was randomly assigned to one of twelve possible JBT orders; across the twelve orders, each face was equally likely to be paired with a more versus less qualified profile.

### ***Self-Report Measures***

Participants completed three self-report items measuring perceived performance, desired performance, and explicit preference. Each item was measured on a 7-point scale ranging from -3 to +3, with 0 indicating neutrality. For perceived performance, participants chose -3 when they were “extremely easier on less physically attractive applicants and tougher on more physically attractive applicants,” and +3 when they were “extremely easier on more physically attractive applicants and tougher on less physically attractive applicants.” For desired performance,

participants chose -3 when they “wanted to be extremely easier on less physically attractive applicants and tougher on more physically attractive applicants” and +3 when they “wanted to be extremely easier on more physically attractive applicants and tougher on less physically attractive applicants.” For explicit preference, participants chose -3 when they “strongly prefer more physically attractive people to less physically attractive people,” and +3 when they “strongly prefer less physically attractive people to more physically attractive people.”

### ***Brief Implicit Association Test***

Participants completed a four-block, good-focal Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009) which measured evaluations toward more versus less physically attractive people. Stimuli for each attractiveness group were two male and two female faces pre-selected from the same images used in the JBT. Responses were analyzed using the *D* scoring algorithm (Nosek et al., 2014), with a higher score indicating more positive implicit associations toward more versus less physically attractive people. See Appendix 2 for BIAT instructions.

## **Results**

In both studies, accuracy on the JBT (accepting more qualified and rejecting less qualified applicants) was above chance (Study 1:  $M = 67.8\%$ ,  $SD = 8.5$ ; Study 2:  $M = 66.9\%$ ,  $SD = 8.8$ ), and levels of sensitivity were above zero (Study 1:  $M = .678$ ,  $SD = .085$ ; Study 2:  $M = .669$ ,  $SD = .088$ ). The average acceptance rate was also close to the recommended 50% (Study 1:  $M = 51.5\%$ ,  $SD = 12.5$ ; Study 2:  $M = 52.4\%$ ,  $SD = 11.8$ ).

### **Criterion Bias and Sensitivity on the JBT**

For both studies, we conducted a paired samples *t*-test in each condition comparing the criterion for more versus less physically attractive applicants. In Study 1, criterion for physically attractive applicants was significantly lower than that for less physically attractive applicants in

all conditions except Implementation Intentions, which did not show any evidence of a criterion bias ( $d = .04$ ). In Study 2, however, criterion for physically attractive applicants was significantly lower than that for less physically attractive applicants in all conditions, meaning bias favoring more physically attractive applicants was present in all conditions. See Table 1 for all means and standard deviations of overall JBT accuracy, sensitivity, and criterion in all conditions in both studies. See Table 2 for all paired samples  $t$ -test statistics.

Next, for both studies, we conducted a series of independent samples  $t$ -tests comparing the degree of criterion bias and overall sensitivity separately for each intervention condition relative to Control. In Study 1, the criterion bias reduction was only significant in the Implementations Intentions condition ( $d = .23$ ), and sensitivity increase was only significant in the Delay condition ( $d = .22$ ). Study 2 replicated these results, with a significant reduction in criterion bias in only the Implementation Intentions condition ( $d = .19$ ) and a significant sensitivity increase in the Delay condition ( $d = .33$ ). See Table 3 for independent-samples test statistics in both studies.

### **Attitudes, Perceived Performance, and Desired Performance**

In a series of exploratory analyses, we then tested whether any of the interventions consistently changed attractiveness attitudes, perceived behavior, or desired behavior. Aside from the exclusion criteria outlined above, we retained all available data (though missing data occurs since participants may have exited the study before completing post-JBT measures or may have decided to skip certain items). Sample sizes and descriptive statistics are presented in Table 4. We conducted a series of independent samples  $t$ -tests comparing the four variables in each intervention condition with the control condition. No consistent results were found across the two studies. See Table 5 for all independent samples  $t$ -test statistics.



## Discussion

Two studies investigated the effectiveness of six different interventions to reduce the magnitude of attractiveness-based discrimination on a hypothetical admissions task. Interventions were evaluated on their ability to either increase sensitivity or reduce attractiveness differences in criterion. First, results consistently replicated the finding of a physical attractiveness bias in decision making using the JBT (Axt, Nguyen & Nosek, 2018), as the criterion for more physically attractive applicants was lower than that for less physically attractive applicants in the Control condition in both studies. Secondly, the results showed that only the Implementation Intentions intervention and the Delay intervention were effective in reducing the magnitude of discrimination compared to the Control condition, but via two different routes. In both Study 1 and 2, the Implementation Intentions intervention significantly reduced criterion bias, and the Delay intervention significantly increased sensitivity.

Consistent with past research, the Implementations Intentions intervention and the Delay intervention were effective in reducing discrimination. The results from this prior work (Axt & Lai, 2019) indicate that the intervention of delay — increasing the amount of time before allowing the participant to make a decision — significantly increased sensitivity, but had no effect on the amount of criterion bias. On the other hand, an awareness-raising intervention that directly warned participants and asked them to avoid using physical attractiveness in their decision-making significantly reduced criterion bias, but had no effect on sensitivity. This awareness intervention broadly resembles the Implementation Intentions intervention in our study, which is the only intervention that directly names the specific driver of bias (i.e., physical attractiveness) and trains participants to avoid it by rehearsing concrete if-then statements.

However, the remaining four interventions had no significant effect in reducing noise or bias. These results are contrary to prior literature, and may be attributed to either a) the altered operationalization of the interventions in our studies or b) the nature of the physical attractiveness bias more broadly.

In the Accountability condition, participants were only alerted that their answers would be evaluated and analyzed by a panel of experts. However, the ineffectiveness of this intervention in the present work may be because the manipulation did not sufficiently heighten accountability. For instance, in a prior study conducted by Kennedy (1993), participants were told that their answer would be evaluated by experts and may be selected for a conference, in which case they would need to explain and justify their decision. The need to provide reasons for one's behavior may be a necessary condition to raise accountability and improve decision-making. Therefore, a future test of the effectiveness of the Accountability intervention may be in alerting participants that they will need to provide a brief explanation of their choice after each selection and asking them to actually provide that justification.

Similarly, in the Confirmation condition, participants only reviewed a 236-word explanation of confirmation bias and a strategy of how to combat it. One potentially important reason why this intervention was ineffective in this work was that the manipulation provided insufficient practice at combatting confirmation bias. In the study reviewed previously (Sellier et al., 2019), participants received multiple interventions to reduce confirmation bias, including a series of exercises eliciting confirmation bias and correction. At the expense of extra time, they were able to better understand, consolidate, and apply the "hypothesis-disconfirming strategy" and inhibit confirmation bias.

Given past research that has provided evidence for the presence of “bias blind spot” (Pronin et al., 2002), where individuals have an easier ability to spot cognitive and motivational biases in others than in themselves, it is possible that simply passively absorbing information about confirmation bias may not be enough to change behavior. Instead, it may be necessary for people to personally experience confirmation bias in order to effectively raise awareness of its existence and motivate the needed regulation strategies. Future studies may then benefit from an enhanced Confirmation intervention that includes several practice questions designed to elicit confirmation bias that can be completed before starting the JBT.

In the Reward condition, participants selected a charity which would receive a \$5 donation if their performance on the JBT were in the top 10<sup>th</sup> percentile. However, this operationalization of reward may have not created enough incentive for participants to invest in greater cognitive resources to inhibit bias and improve accuracy. There is a considerable body of literature demonstrating the causal relationship between performance-contingent financial incentive and better decision-making (Stone & Ziebert, 1995; Tosi et al., 1997), and their financial incentives were in the form of direct monetary transfer. It may be that the ineffectiveness of the Reward intervention in our study was due to a weaker financial incentive relative to past work. To address this issue, a subsequent study could test whether informing participants that they themselves would receive a \$10 payment for superior performance could significantly reduce discrimination.

Finally, in the Objectivity condition, participants were asked to evaluate all applicants objectively and write down the information they will use when evaluating and why they are relevant. However, because the applicant profiles in this version of the JBT contain both male- and female-typical faces, participants might have tried to avoid gender bias, the lack of

specifying *what* participants should be objective about may have inadvertently led them to focus on applicant gender and not physical attractiveness. In a prior study (Uhlmann & Cohen, 2005), the profiles differed only in terms of names, indicating a male and a female respectively. In the present study, the more salient dimension of gender might have served as a potential distractor, causing participants to focus too much on inhibiting gender bias and not on inhibiting biases based on physical attractiveness. Like past research has shown (Axt et al., 2019), asking participants to avoid biases in general when there are multiple sources of bias present will have little effect in reducing either bias. Therefore, to more clearly evaluate the effect of the Objectivity condition, we may need to design a version of the JBT that only contains profiles from one gender.

The operationalizations used here have several notable departures from prior uses of each intervention. However, an alternative reason for why several interventions failed to reduce discrimination is the nature of the physical attractiveness bias. It is possible that biases based on physical attractiveness are more subtle and more resistant to change than biases based on, for example, race or gender. Relatedly, there may be a large number of people who sincerely did not think they used physical attractiveness in their judgments (Axt et al., 2018). In this case, even if the participants were committed to being generally objective, this enhanced objectivity may have been ineffective at changing behavior if they did not realize the influence of attractiveness on their judgments. From this perspective, it is possible that the above-listed conditions are in fact adequate operationalizations of each bias-reducing approach, but such interventions just did not effectively translate into reduced *attractiveness* bias in this context. This perspective may in turn explain why only the Implementation Intentions intervention effectively reduced bias, as it was the only intervention that explicitly named the physical attractiveness bias. One clear next step is

to explore directly whether mentioning physical attractiveness biases specifically in the instructions of each interventions could inhibit bias (Axt et al., 2019), even when using very similar operationalizations to what was deployed here.

### **Limitations**

On the whole, Study 1 and 2 offer highly consistent results. The minor inconsistencies could potentially be explained mainly by different samples. The participants in Study 1 were recruited from Project Implicit, who are, in general, younger, more highly educated, more concentrated in Europe and North America, and more knowledgeable about bias and bias reduction. On the other hand, the participants in Study 2 were recruited from Prolific and received \$1 after completing the study. The Prolific participants completed the task for monetary incentive and, in general, were likely less knowledgeable about bias and bias reduction compared to Project Implicit participants. Greater knowledge about bias or motivation to address it may explain why both mean JBT accuracy rate across interventions was higher in Study 1 compared to Study 2.

A major limitation of the present studies is that the JBT is designed to be short and quick to complete, which means that only small, one-shot interventions could be employed. Many studies on reducing discrimination (Aboud & Levy, 2000; Becker et al., 2014; Gronholm et al., 2017) have tested interventions that are more intensive and extensive, requiring more time and resources. These interventions are presumably more effective and long-lasting than the interventions used in the present studies. For example, an intensive intervention employed to reduce mental health discrimination lasted several hours and provided education via group discussions and films (Gronholm et al., 2017). Similarly, an extensive intervention in adolescents required repeated training in various social-cognitive skills in an effort to reduce prejudice and

discrimination, (Aboud & Levy, 2000). In the context of this work, it's possible that a more intensive intervention to reduce attractiveness-based discrimination, such as watching a film with a less physically attractive protagonist and then holding a discussion session about the material, may translate into more effective behavior change.

Finally, the present studies do not provide a longitudinal analysis of the effect of the six interventions in reducing discrimination. Due to the nature of online platforms like Project Implicit and Prolific, it is difficult to conduct longitudinal studies on a large scale. However, it is important to test the longevity of bias-reducing interventions (Lai et al., 2016). Some interventions, like delay or implementation intentions, may have limited long-term effects and not translate into future situations where decision-makers are not explicitly reminded to slow down or develop a strategy for regulating bias. As a result, it would be productive to explore whether if rehearsing distraction-inhibiting implementation intentions could have a significant effect in reducing discrimination one day and one week after the initial intervention.

### **Future Directions**

Other than altering the operationalization of the interventions and adding a longitudinal component to the study, two other directions are worth exploring in the future. First, we could further this work by examining the effect of combining different interventions. The study conducted by Sellier et al. (2019) created an intervention bundle which, together, had a significant effect in reducing confirmation bias. A previous study (Axt & Lai, 2019) has shown that asking participants to avoid favoring more physically attractive applicants *and* requiring them to take more time before making a decision had a significant effect in both increasing sensitivity and reducing attractiveness-based biases in response criterion. In the case of the present studies, combining two interventions which, on their own, had no significant impact on

reducing discrimination could potentially reduce bias and/or increase sensitivity. For instance, it would be interesting to combine the Objectivity and the Confirmation condition, as they complement each other well. By committing to objective standards and preventing confirmation bias, participants may be adequately motivated to change their behavior, resulting in greater accuracy and/or decreased bias.

Another important direction for this work concerns the issue of intersectionality. Most research, including the present studies, has focused on a single category of bias at a time, such as physical appearance. However, in reality, people's identities are composed of multiple aspects at once, including gender, race, age, ability, class, education, religion, and other group affiliations. A previous study (Axt et al., 2019) has shown that when multiple sources of bias are present, such as a profile containing both a photo (eliciting physical attractiveness bias) and university affiliation (eliciting in-group favouritism bias), identifying and asking participants to avoid one source of bias significantly reduced that specific bias, but not biases for the unmentioned category. Furthermore, alerting participants to bias in general and not providing any mention of the social information that was driving bias did not consistently change judgment. Given how important intersectionality is in real-world decision-making and discrimination (Garnett et al., 2013; Fielden & Davidson, 2012), it is necessary to develop interventions, either by adapting those used here or developing novel ones, that can impact multiple biases operating simultaneously.

## **Conclusion**

Two studies investigated the effectiveness of six interventions in reducing discrimination based on physical attractiveness. Results found that rehearsing distraction-inhibiting if-then statements and requiring a delay in response significantly reduced the magnitude of

discrimination, through decreasing bias and increasing accuracy respectively. This work is the largest comparative assessment of strategies to reduce biased judgment, and the results provide an important foundation to future work on the study of discrimination reduction.



### Reference

- About, F. E., & Levy, S. R. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *"The Claremont Symposium on Applied Social Psychology" Reducing prejudice and discrimination* (p. 269–293). Lawrence Erlbaum Associates Publishers.
- Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2007). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology*. <https://doi.org/10.1111/j.1469-8986.2007.00609.x>
- Ariely, D., & Zakay, D. (2001). A timely account of the role of duration in decision making. *Acta Psychologica*, 108(2), 187–207. [https://doi.org/10.1016/s0001-6918\(01\)00034-8](https://doi.org/10.1016/s0001-6918(01)00034-8)
- Axt, J.R., Casola, G.M. & Nosek, B.A (2019). Reducing social judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin*, 45, 1232-1251. <https://doi.org/10.31234/osf.io/ngxks>
- Axt, J.R. & Lai, C.K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology*, 117, 26-49. <https://doi.org/10.1037/pspa0000153>
- Axt, J. R., Nguyen, H., & Nosek, B. A. (2018). The Judgment Bias Task: A flexible method for assessing individual differences in social judgment biases. *Journal of Experimental Social Psychology*, 76, 337-355. <https://doi.org/10.1016/j.jesp.2018.02.011>
- Becker, J. C., Zawadzki, M. J., & Shields, S. A. (2014). Confronting and Reducing Sexism: A Call for Research on Intervention. *Journal of Social Issues*, 70(4), 603–614. <https://doi.org/10.1111/josi.12081>

- Commisso, M., & Finkelstein, L. (2012). Physical Attractiveness Bias in Employee Termination. *Journal of Applied Social Psychology, 42*(12), 2968–2987. <https://doi.org/10.1111/j.1559-1816.2012.00970.x>
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin, 87*(3), 546–563. <https://doi.org/10.1037/0033-2909.87.3.546>
- Devine, P. G., & Elliot, A. J. (1995). Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited. *Personality and Social Psychology Bulletin, 21*(11), 1139–1150. <https://doi.org/10.1177/01461672952111002>
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science, 11*(4), 315–319. <https://doi.org/10.1111/1467-9280.00262>
- Fielden, S., & Davidson, M. J. (2012). BAME women business owners: how intersectionality affects discrimination and social support. *Gender in Management: An International Journal, 27*(8), 559–581. <https://doi.org/10.1108/17542411211279733>
- Fiske, S. T. (1998). *Stereotyping, prejudice, and discrimination*. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (p. 357–411). McGraw-Hill.
- Fitzgerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychology, 7*(1). <https://doi.org/10.1186/s40359-019-0299-7>
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*(3), 522–559. <https://doi.org/10.1037/pspa0000160>

- Garnett, B. R., Masyn, K. E., Austin, S. B., Miller, M., Williams, D. R., & Viswanath, K. (2013). The Intersectionality of Discrimination Attributes and Bullying Among Youth: An Applied Latent Class Analysis. *Journal of Youth and Adolescence*, 43(8), 1225–1239. <https://doi.org/10.1007/s10964-013-0073-8>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.
- 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(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Gronholm, P. C., Henderson, C., Deb, T., & Thornicroft, G. (2017). Interventions to reduce discrimination and stigma: The state of the art. *Social Psychiatry and Psychiatric Epidemiology*, 52(3), 249–258. <https://doi.org/10.1007/s00127-017-1341-9>
- Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The Effects Of Physical Attractiveness On Job-Related Outcomes: A Meta-Analysis Of Experimental Studies. *Personnel Psychology*, 56(2), 431–462. <https://doi.org/10.1111/j.1744-6570.2003.tb00157.x>
- Kattari, S. K., Whitfield, D. L., Walls, N. E., Langenderfer-Magruder, L., & Ramos, D. (2016). Policing Gender Through Housing and Employment Discrimination: Comparison of Discrimination Experiences of Transgender and Cisgender LGBTQ Individuals. *Journal of the Society for Social Work and Research*, 7(3), 427–447. <https://doi.org/10.1086/686920>
- Kennedy, J. (1993). Debiasing Audit Judgment with Accountability: A Framework and Experimental Results. *Journal of Accounting Research*, 31(2), 231. <https://doi.org/10.2307/2491272>
- Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., ... Nosek, B. A. (2014). Reducing implicit RACIAL Preferences: I. a comparative investigation of 17

- interventions. *Journal of Experimental Psychology: General*, 143(4), 1765–1785. <https://doi.org/10.1037/a0036260>
- Lee, Y. H. (2019). Strengths and Limitations of Meta-Analysis. *The Korean Journal of Medicine*, 94(5), 391–395. <https://doi.org/10.3904/kjm.2019.94.5.391>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255-275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, 36, 512-523. <https://doi.org/10.1177/0146167210362789>
- 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. <https://doi.org/10.1037/0022-3514.65.3.469>
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479. <https://doi.org/10.1073/pnas.1211286109>
- Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the brief Implicit Association Test: recommended scoring procedures. *PloS one*, 9(12), e110938. <https://doi.org/10.1371/journal.pone.0110938>
- 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.), *Frontiers of social psychology. Social psychology and the unconscious: The automaticity of higher mental processes* (p. 265–292). Psychology Press.

- Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What Works? A Review and Assessment of Research and Practice. *Annual Review of Psychology*, 60(1), 339–367. <https://doi.org/10.1146/annurev.psych.60.110707.163607>
- 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(3), 369–381. <https://doi.org/10.1177/0146167202286008>
- Romanelli, M., & Hudson, K. D. (2017). Individual and systemic barriers to health care: Perspectives of lesbian, gay, bisexual, and transgender adults. *American Journal of Orthopsychiatry*, 87(6), 714–728. <https://doi.org/10.1037/ort0000306>
- Rooth, D.-O. (2009). Obesity, Attractiveness, and Differential Treatment in Hiring: A Field Experiment. *Journal of Human Resources*, 44(3), 710–735. <https://doi.org/10.1353/jhr.2009.0027>
- Ross, S. L., & Turner, M. A. (2005). Housing Discrimination in Metropolitan America: Explaining Changes between 1989 and 2000. *Social Problems*, 52(2), 152–180. <https://doi.org/10.1525/sp.2005.52.2.152>
- Ross, S. L., & Yinger, J. (2003). *The color of credit: mortgage discrimination, research methodology, and fair-lending enforcement lending*. MIT Press.
- Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing Training Improves Decision Making in the Field. *Psychological Science*, 30(9), 1371–1379. <https://doi.org/10.1177/0956797619861429>
- Shaffer, D. R., Crepaz, N., & Sun, C.-R. (2000). Physical Attractiveness Stereotyping in Cross-Cultural Perspective: Similarities and Differences between Americans and Taiwanese.

- Journal of Cross-Cultural Psychology*, 31(5), 557–582. <https://doi.org/10.1177/0022022100031005002>
- Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, 26, 278-292. <https://doi.org/10.1086/209563>
- Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, 56(4), 283–294. <https://doi.org/10.1027/1618-3169.56.4.283>
- Stone, D. N., & Ziebart, D. A. (1995). A Model of Financial Incentive Effects in Decision Making. *Organizational Behavior and Human Decision Processes*, 61(3), 250–261. <https://doi.org/10.1006/obhd.1995.1020>
- Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to implicit association tests and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (p. 117–139). The Guilford Press.
- Thorat, S., & Attewell, P. (2007). The Legacy of Social Exclusion: A Correspondence Study of Job Discrimination in India. *Economic and Political Weekly*, 42(41), 4141-4145. Retrieved December 13, 2020, from <http://www.jstor.org/stable/40276548>
- Tosi, H. L., Katz, J. P., & Gomez-Mejia, L. R. (1997). Disaggregating The Agency Contract: The Effects Of Monitoring, Incentive Alignment, And Term In Office On Agent Decision Making. *Academy of Management Journal*, 40(3), 584–602. <https://doi.org/10.5465/257054>

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*, 474-480. <https://doi.org/10.1111/j.0956-7976.2005.01559.x>

## **Appendix 1: Instructions for the 7 Conditions**

### **Control Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

### **Accountability Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

It is important to note that your decisions on the task will be reviewed by a panel of researchers that are experienced in evaluating students. They will analyze your performance in terms of accurately accepting more qualified applicants and rejecting less qualified applicants.

### **Confirmation Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

Research has shown that in order to be accurate on the task, it is important to reduce one's susceptibility to something called 'confirmation bias'. Confirmation bias occurs when someone selectively searches for information that validates their hypothesis and ignores information that may be inconsistent with their beliefs. As a result, confirmation bias means that people will ignore other important information that might indicate that their hypothesis is, in fact, incorrect. For example, an individual with a fear of flying in airplanes is more likely to search the internet



for news reports about airplane accidents while ignoring robust data suggesting that airplanes are quite safe.

Confirmation bias can affect an individual's decisions on the academic selection task. For example, under confirmation bias, if one believes a student to be qualified, they might focus on the student's high interview score while ignoring their low GPA and low recommendation letter strength. Alternatively, if one believes a student to be unqualified, they might focus on the student's low GPA while ignoring their high interview score and recommendation letter strength.

In order to overcome confirmation bias, research has shown it to be helpful to adopt a 'hypothesis disconfirming' strategy. You can adopt this strategy by looking for information that suggests a student may be unqualified, rather than only looking for information that suggests a student may be qualified. Adopting this strategy will make it easier to eliminate students that are unqualified.

**Delay Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

Past research suggests that people may be more accurate at evaluating applicants if they can slow down and spend more time reflecting on each decision. To help you do so, there will be a four-second delay between when the application is first presented and when you will be able to make an accept or reject decision. You can use those four seconds to think more about your decision.

**Implementation Intentions Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

In order to select the most qualified applicants, you should be careful to not let irrelevant information affect your decisions. In order to help you achieve this, research has shown it will be helpful for you to adopt the following strategy: *If I see a student's application, then I will ignore their face.*

Please mentally repeat this strategy three times using inner speech. When you are comfortable recalling it, go on to the next page.

Type out the strategy you learned on the previous page in the box below.

### **Objectivity Condition Instruction**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

Studies have shown that although people want to be accurate on the task, they often show behavior that is inconsistent with this goal.

In order to meet this goal of being accurate, it may be helpful for you to adopt an objective mindset. This can be accomplished by deciding on a set of parameters you will use to evaluate students **before** beginning the task. Committing to such a strategy beforehand is important because it could prevent you from being distracted by irrelevant information during the task.

Before you begin the task, we want to know more about how you will objectively evaluate each application. Please answer the following questions in the space provided.

What information will you consider when evaluating the applicants?

Why is the information you listed in the previous question relevant to your evaluation?

**Reward Condition Instruction:**

In this study you will be asked to imagine you are on the selection committee for an academic honor society. During the task, you will decide whether to accept or reject a number of applicants.

We are interested in your ability to accept the more qualified applicants and reject the less qualified applicants. To motivate you to perform well, participants who are in the top 10% for accuracy will have a \$5 donation made to a charity of their choice.

Below is a list of fifteen charities that have earned an A rating or higher from the independent website [charitywatch.org](http://charitywatch.org): Bowery Residents Committee (BRC), Brain & Behavior Research Foundation Breast Cancer Research Foundation, Childrens Defense Fund, Compassion International, Elizabeth Glaser Pediatric AIDS Foundation, Farm Aid, Fisher House Foundation, Food and Water Watch, Goodwill Industries International (American National Office), Helen Keller International, International Peace Institute, Lupus Research Alliance, Scholarship America, Wildlife Conservation Society.



Please select the charity that you would like to receive a donation if you are in the top 10% in terms of accuracy on the selection task.

## Appendix 2: BIAT Instructions

### Brief Implicit Association Test

Next, you will use the 'E' and 'I' computer keys to categorize words or images into groups as fast as you can.

These are the four groups and the words or images that belong to each:

Category	Stimuli
More Attractive People	
Less Attractive People	
Good Words	LOVE, PLEASANT, GREAT, WONDERFUL
Bad Words	HATE, UNPLEASANT, AWFUL, TERRIBLE

This portion of the study takes 3 minutes on average. When you are done, you will receive information about the purpose of the study and feedback on how you performed. Press the button below when you are ready to begin.

[Continue](#)

"E" for all else **More Attractive People** "I" if item belongs



and  
**Good Words**  
 Love, Pleasant, Great, Wonderful  
 Part 1 of 4

Put a right finger on the I key for items that belong to the category **Good Words**, and for items that belong to the category **More Attractive People**. Put a left finger on the E key for items that do not belong to these categories.

Items will appear one at a time.

If you make a mistake, a red X will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

"E" for all else **Less Attractive People** "I" if item belongs



and  
**Good Words**  
 Love, Pleasant, Great, Wonderful  
 Part 2 of 4

Put a right finger on the I key for items that belong to the category **Good Words**, and for items that belong to the category **Less Attractive People**. Put a left finger on the E key for items that do not belong to these categories.

Items will appear one at a time.

If you make a mistake, a red X will appear. Press the other key to continue.

Press the **space bar** when you are ready to start.

"E" for all else **More Attractive People** "I" if item belongs

and  
**Good Words**



If you make a mistake, a red X will appear. Press the other key to continue.

**Table 1: Means and Standard Deviations for Overall JBT Accuracy, Sensitivity, and Criterion for each condition**

<b>Condition</b>	<b>JBT Accuracy Mean (SD)</b>	<b>Sensitivity Mean (SD)</b>	<b>More Attractive Criterion Mean (SD)</b>	<b>Less Attractive Criterion Mean (SD)</b>
<b>Study 1</b>				
Control (N = 545)	67.45% (8.54)	1.01 (.54)	.12 (.45)	.01 (.46)
Accountability (N = 554)	67.30% (8.85)	1.01 (.57)	.10 (.45)	-.04 (.49)
Confirmation (N=548)	68.10% (8.49)	1.05 (.55)	.11 (.44)	.02 (.48)
Delay (N = 530)	69.36% (7.73)	1.13 (.52)	.16 (.44)	.05 (.44)
Implementation (N = 507)	67.60% (8.91)	1.01 (.57)	.05 (.42)	.03 (.44)
Objectivity (N = 500)	67.11% (8.75)	1.00 (.56)	.04 (.44)	-.04 (.49)
Reward (N = 538)	67.69% (8.33)	1.04 (.54)	.07 (.46)	-.05 (.48)
<b>Study 2</b>				
Control (N = 343)	66.29% (7.74)	.92 (.47)	.16 (.40)	.004 (.47)
Accountability (N = 395)	65.87% (9.37)	.90 (.56)	.16 (.42)	-.001 (.46)
Confirmation (N = 367)	66.48% (9.95)	.95 (.62)	.11 (.40)	.003 (.45)
Delay (N = 402)	69.08% (7.77)	1.08 (.49)	.21 (.38)	.02 (.40)
Implementation (N = 338)	66.37% (8.49)	.94 (.53)	.10 (.46)	.03 (.44)
Objectivity (N = 439)	66.72% (8.23)	.96 (.51)	.09 (.45)	-.02 (.45)
Reward (N = 375)	66.93% (9.33)	.97 (.57)	.12 (.43)	-.003 (.43)
Note: Higher criterion value denotes greater leniency Higher criterion bias denotes more leniency towards more physically attractive				

**Table 2: Criterion Bias for each condition**

<b>Condition</b>	<b>Criterion Bias</b>
<b>Study 1</b>	
Control	$t(544) = 5.89, p < .001, d = .25 [.17, .34]**$
Accountability	$t(553) = 6.62, p < .001, d = .22 [.20, .37]**$
Confirmation	$t(547) = 4.46, p < .001, d = .19 [.11, .28]**$
Delay	$t(529) = 6.40, p < .001, d = .28 [.19, .37]**$
Implementation	$t(506) = 1.00, p = .317, d = .04 [-.04, .13]$
Objectivity	$t(499) = 4.20, p < .001, d = .19 [.10, .28]**$
Reward	$t(537) = 6.67, p < .001, d = .29 [.20, .37]**$
<b>Study 2</b>	
Control	$t(342) = 6.33, p < .001, d = .25 [.17, .34]**$
Accountability	$t(394) = 6.88, p < .001, d = .35 [.25, .45]**$
Confirmation	$t(366) = 5.33, p < .001, d = .28 [.17, .38]**$
Delay	$t(401) = 8.59, p < .001, d = .43 [.33, .53]**$
Implementation	$t(337) = 3.32, p = .001, d = .18 [.07, .29]**$
Objectivity	$t(348) = 4.98, p < .001, d = .27 [.16, .37]**$
Reward	$t(374) = 5.59, p < .001, d = .29 [.19, .39]**$
Note: Criterion Bias = Within-subjects test comparing criterion values. ** denotes statistical significance	

**Table 3: Independent Samples t-test between Control and 6 Intervention Conditions on Criterion Bias and Sensitivity**

Intervention Condition	Criterion Bias	Sensitivity
<b>Study 1</b>		
Accountability	t(1097) = .92, p = .358, d = .06 [-.06, .17]	t(1097) = -.23, p = .816, d = -.01 [-.13, .10]
Confirmation	t(1091) = -1.08, p = .282, d = -.07 [-.18, .05]	t(1091) = 1.22, p = .222, d = .07 [-.05, .19]
Delay	t(1074) = .33, p = .745, d = .02 [-.10, .14]	t(1074) = 3.60, p < .001, d = .220 [.10, .34]**
Implementation	t(1050) = -3.64, p < .001, d = -.23 [-.35, -.10]**	t(1050) = .12, p = .903, d = .01 [-.11, .13]
Objectivity	t(1043) = -1.05, p = .296, d = -.07 [-.19, .06]	t(1043) = -.43, p = .673, d = -.03 [-.15, .10]
Reward	t(1081) = .47, p = .639, d = .03 [-.09, .15]	t(1081) = .66, p = .508, d = .04 [-.08, .16]
<b>Study 2</b>		
Accountability	t(736) = .05, p = .961, d = .004 [-.14, .15]	t(736) = -.37, p = .715, d = -.03 [-.17, .12]
Confirmation	t(708) = -1.53, p = .126, d = -.12 [-.26, .03]	t(708) = .66, p = .508, d = .05 [-.10, .20]
Delay	t(743) = .95, p = .342, d = .07 [-.07, .21]	t(743) = 4.55, p < .001, d = .33 [.19, .48]**
Implementation	t(679) = -2.49, p = .013, d = -.19 [-.34, -.04]**	t(679) = .55, p = .581, d = .04 [-.11, .19]
Objectivity	t(690) = -1.25, p = .212, d = -.10 [-.24, .05]	t(690) = 1.08, p = .282, d = .08 [-.07, .23]
Reward	t(716) = -.88, p = .381, d = -.07 [-.21, .08]	t(716) = 1.25, p = .212, d = .09 [-.05, .23]
<p>Note: Criterion Bias = Between-subjects test comparing criterion values of the control and the specified intervention  Sensitivity = Between-subjects test comparing sensitivity values of the control and the specified intervention  ** denotes statistical significance  Positive t statistic denotes increase in criterion bias and sensitivity, vice versa.</p>		

**Table 4: Sample Size, Mean, and Standard Deviation for Implicit Attitude, Explicit Attitude, Perceived Performance, and Desired Performance in each condition**

Condition	Implicit Attitude	Explicit Attitude	Perceived Performance	Desired Performance
<b>Study 1</b>				
Control	N = 500, M = .65, SD = .47	N = 494, M = 3.37, SD = .82	N = 496, M = 4.12, SD = .59	N = 495, M = 4.00, SD = .39
Accountability	N = 497, M = .66, SD = .52	N = 494, M = 3.39, SD = .89	N = 500, M = 4.10, SD = .67	N = 491, M = 3.97, SD = .36
Confirmation	N = 496, M = .74, SD = .47	N = 485, M = 3.28, SD = .85	N = 488, M = 4.06, SD = .55	N = 479, M = 3.99, SD = .35
Delay	N = 489, M = .72, SD = .51	N = 483, M = 3.37, SD = .89	N = 488, M = 4.10, SD = .58	N = 485, M = 3.99, SD = .39
Implementation	N = 470, M = .64, SD = .51	N = 460, M = 3.33, SD = .87	N = 461, M = 4.06, SD = .52	N = 455, M = 4.00, SD = .33
Objectivity	N = 447, M = .68, SD = .48	N = 436, M = 3.35, SD = .86	N = 438, M = 4.11, SD = .52	N = 433, M = 4.00, SD = .34
Reward	N = 492, M = .69, SD = .49	N = 483, M = 3.35, SD = .90	N = 490, M = 4.06, SD = .54	N = 478, M = 3.97, SD = .34
<b>Study 2</b>				
Control		N = 340, M = 3.26, SD = 1.00	N = 340, M = 4.11, SD = .82	N = 338, M = 4.04, SD = .67
Accountability		N = 392, M = 3.22, SD = 1.04	N = 394, M = 4.08, SD = .90	N = 391, M = 3.95, SD = .66
Confirmation		N = 364, M = 3.16, SD = 1.02	N = 366, M = 3.97, SD = .80	N = 360, M = 3.93, SD = .69
Delay	N/A	N = 397, M = 3.24, SD = .97	N = 396, M = 4.11, SD = .79	N = 398, M = 4.00, SD = .65
Implementation		N = 333, M = 3.17, SD = 1.01	N = 333, M = 3.97, SD = .79	N = 328, M = 4.03, SD = .55
Objectivity		N = 345, M = 3.29, SD = .99	N = 347, M = 3.99, SD = .76	N = 342, M = 3.98, SD = .65
Reward		N = 372, M = 3.28, SD = .94	N = 374, M = 4.06, SD = .75	N = 368, M = 3.93, SD = .63
<p>Note: For implicit attitude, perceived performance, and desired performance, higher mean denotes more positive association with / perceived favoritism for / desired favoritism for more physically attractive people.</p> <p>For explicit attitude, lower mean denotes more positive association with / perceived favoritism for / desired favoritism for more physically attractive people.</p>				



**Table 5: Independent Samples t-test between Control and 6 Intervention Conditions on Implicit Attitude, Explicit Attitude, Perceived Performance, and Desired Performance**

Intervention Condition	Implicit Attitude	Explicit Attitude	Perceived Performance	Desired Performance
<b>Study 1</b>				
Accountability	t(995) = .15, p = .884, d = .01 [-.12, .13]	t(983) = -.29, p = .769, d = -.02 [-.14, .11]	t(994) = -.47, p = .636, d = -.03 [-.15, .09]	t(984) = -1.11, p = .268, d = -.07 [-.20, .05]
Confirmation	t(994) = 3.01, p = .003, d = .19 [.07, .32]**	t(974) = 1.73, p = .084, d = .11 [-.02, .24]	t(982) = -1.75, p = .081, d = -.11 [-.24, .01]	t(972) = -.62, p = .538, d = -.04 [-.17, .09]
Delay	t(987) = 2.06, p = .039, d = .13 [.01, .26]**	t(972) = .08, p = .939, d = .01 [-.12, .13]	t(982) = -.39, p = .698, d = -.03 [-.15, .10]	t(978) = -.33, p = .740, d = -.02 [-.15, .10]
Implementation	t(968) = -.304, p = .761, d = -.02 [-.15, .11]	t(949) = .85, p = .394, d = .06 [-.07, .18]	t(955) = -1.74, p = .082, d = -.11 [-.24, .01]	t(948) = -.00, p = 1.000, d = -.00 [-.13, .13]
Objectivity	t(945) = .929, p = .353, d = .06 [-.07, .19]	t(925) = .44, p = .661, d = .03 [-.10, .16]	t(932) = -.19, p = .846, d = -.01 [-.14, .12]	t(926) = -.00, p = 1.000, d = -.00 [-.13, .13]
Reward	t(990) = 1.20, p = .232, d = .08 [-.05, .20]	t(972) = .42, p = .678, d = .03 [-.01, .15]	t(984) = -1.55, p = .121, d = -.10 [-.22, .03]	t(971) = -1.07, p = .284, d = -.07 [-.19, .06]
<b>Study 2</b>				
Accountability		t(730) = .53, p = .596, d = .04 [-.11, .19]	t(732) = -.59, p = .553, d = -.04 [-.19, .10]	t(727) = -1.93, p = .054, d = -.14 [-.29, .00]
Confirmation		t(702) = 1.31, p = .190, d = .10 [-.05, .25]	t(704) = -2.27, p = .023, d = -.17 [-.32, .02]**	t(696) = -2.26, p = .024, d = -.17 [-.32, .02]**
Delay	N/A	t(735) = .28, p = .770, d = .02 [-.12, .17]	t(734) = -.05, p = .958, d = -.00 [-.15, .14]	t(734) = -.96, p = .337, d = -.07 [-.21, .07]
Implementation		t(671) = 1.28, p = .199, d = .10 [-.05, .25]	t(671) = -2.28, p = .023, d = -.18 [-.33, .03]**	t(664) = -.29, p = .771, d = -.02 [-.17, .13]
Objectivity		t(683) = -.29, p = .769, d = -.02 [-.17, .13]	t(685) = -2.09, p = .037, d = -.16 [-.31, .01]**	t(678) = -1.28, p = .201, d = -.10 [-.25, .05]
Reward		t(710) = -.28, p = .780, d = -.02 [-.17, .13]	t(712) = -.90, p = .368, d = -.07 [-.21, .08]	t(704) = -2.29, p = .022, d = -.17 [-.32, .03]**

Note: Implicit Attitude = Between-subjects test comparing BIAT D score of the control and the specified intervention  
 Explicit Attitude = Between-subjects test comparing explicit attitude self-report values of the control and that of the specified intervention  
 Perceived Performance = Between-subjects test comparing perceived performance self-report values of the control and that of the specified intervention  
 Desired Performance = Between-subjects test comparing desired performance self-report values of the control and that of the specified intervention  
 \*\* denotes statistical significance  
 Positive t statistic denotes more positive association with / preference for / perceived favoritism for / desired favoritism for more physically attractive people compared to control condition