Using a Consider the Opposite Intervention to Reduce Attractiveness Bias on the Judgment Bias Task

Angela Gao

Department of Psychology, McGill University

PSYC 498D1/D2: Senior Honours Research

Dr. Jordan Axt

April 2024

Abstract

Attractiveness bias is a pervasive form of bias that can influence decision-making in hiring. Certain debiasing interventions, classically used to reduce judgment errors, have previously been shown to also reduce intergroup bias. The current study furthers this line of research by investigating a "consider the opposite" intervention to reduce attractiveness bias. We hypothesized that the "consider the opposite" intervention (i.e., prompting participants to think why someone may reach the opposite decision) would increase accuracy and reduce bias on the Judgement Bias Task, where participants (n = 480) choose applicants varying in competence and attractiveness for an honours society. An exploratory analysis was also done to investigate response differences between White and non-White participants. Independent-samples t-tests revealed that "consider the opposite" did not reliably reduce bias or accuracy. In addition, "consider the opposite" reliably reduced bias among White participants but not non-White participants. Overall, this study suggests that attractiveness bias may function too differently from judgment biases for "consider the opposite" to be effective. However, it is notable that the effect size for bias reduction was marginally significant, and replication in a larger sample is required to be confident in our results. Exploratory analyses also reveal that there may be an ingroup effect affecting the "consider the opposite" intervention. Future studies should replicate the current study in a larger and more diverse sample to better mimic real-life hiring conditions.

Introduction

Intergroup bias is a preference towards certain groups over others (Tajfel et al., 1979). These biases occur comparatively automatically, and prior research has illustrated that such biases can exist outside of conscious awareness (e.g., Devine, 1989). Group formation for intergroup biases is often based on demographic characteristics such as race and sex; however, other social dimensions can also elicit bias. Attractiveness bias, a bias favoring physically attractive people, is an example of bias that is not necessarily based on a demographic group or ingroup status. Despite being less studied, the pervasive nature of attractiveness bias has translated to advantages of attractiveness in several aspects of life, including hiring outcomes (Hosoda, 2006; Marlowe et al., 1996), perception of academic performance (Talamas et al., 2016), financial negotiations (Maestripieri et al., 2016), and perceptions of competence (Eagly et al., 1991).

One current framework for the basis of favouritism based on physical attractiveness comes from implicit personality theory, which states that individuals automatically assume someone has a set of commonly associated characteristics based on their other traits. In the context of attractiveness, this perspective argues that people automatically associate various positive traits, such as competency, intelligence, and influence, with physical attractiveness (Eagly et al., 1991). This phenomenon is also described in past work on the "halo effect," which refers to the tendency for positive impressions in one area to influence judgments about the same person in an unrelated area (Wilson & Brekke, 1994). Indeed, evidence supporting this relationship between physical attractiveness and positive personality traits has been found in many studies. For instance, Talamas and colleagues (2016) investigated the relationship between physical attractiveness and perceived academic performance, asking participants to rate a series

of 100 faces on perceived attractiveness, intelligence, conscientiousness, and academic performance. While results found no relationship between perceived attractiveness and actual academic performance (r = .03), there were positive correlations between perceived attractiveness and both perceived intelligence (r = .81) and conscientiousness (r = .74). Given that such traits are deemed as important during hiring and academic contexts (Barrick & Mount, 2012), the prevalence of attractiveness bias in many consequential settings is perhaps not surprising.

Considering the variety of scenarios where attractiveness has been shown to influence judgment, investigating strategies to reduce bias warrants investigation. Evidence-based strategies to reduce bias should be tested and applied to real-world settings, especially considering prior work suggesting that people lack insight into what interventions most impact discriminatory behavior. For instance, a collection of studies done by Axt and colleagues (2023) investigated disparities between the predicted and actual effectiveness of several classic bias reduction strategies when applied to a decision-making task known to consistently produce biases favouring more physically attractive people. Specifically, the researchers investigated six strategies: implementation strategies (i.e., if-then strategies for planning unbiased behaviors; Mendoza et al., 2010), time delay (i.e., forcing the participant to wait before making a decision/performing an action; Axt & Lai, 2019), increasing accountability (Lerner & Tetlock, 1999), educating about confirmation bias (i.e. warning about the tendency to interpret information in a way that is consistent with prior beliefs; Sellier et al., 2019), asking participants to be objective in their decision (Fischhoff, 1982), and offering rewards for remaining accurate and unbiased (Stone & Ziebart, 1995).

Of the six interventions, only implementation strategies and time delay produced reliable results on actual behavior (i.e., reducing the magnitude of attractiveness-based discrimination on the hypothetical decision-making task). However, there was little overlap between the strategies that lay participants *thought* would work and the strategies that actually worked. Less than 1% of participants correctly identified the combination of strategies that did versus did not effectively reduce discrimination in the judgment task (Axt et al., 2023). Given the limited evidence that individuals can effectively generate or identify bias-reducing interventions, researchers should continue to test a variety of interventions to determine evidence-based interventions that can be applied to real-life scenarios.

Besides the discrimination-reducing interventions tested in Axt et al. (2023), other possible debiasing interventions have been shown to reduce judgment errors in separate outcomes and contexts. Here, judgment errors refer to mistakes caused by non-relevant information, which are conceptually different from errors caused by intergroup bias against social groups (Benjamin, 2019). Some commonly known judgment errors include the gambler's fallacy (the belief that an event is more likely to happen based on previous events; Terrell, 1994) and the anchoring bias (the tendency to rely too heavily on an initial impression or initially provided information; Tversky & Kahneman, 1974).

These types of judgment biases may share some similar mechanisms as biases in intergroup behavior, in that both depend on an irrational influence of certain information on judgment. However, debiasing interventions and intergroup bias interventions are often studied as two distinct collections of strategies. As a result, many popular debiasing interventions have not been applied to investigate their effects on intergroup bias. Given the possible similarities between intergroup and judgment bias, interventions from one domain (i.e., biases in research on

judgment and decision-making) may also demonstrate effectiveness in other domains (i.e., biases in research on intergroup discrimination).

Consider the Opposite

One well-known intervention in the debiasing literature is "consider the opposite." First introduced by Lord and colleagues in 1984, the consider the opposite strategy works by providing instructions to consider various hypothetical and opposite possibilities to what participants would typically focus on (Lord et al., 1984). For example, Mussweiler and colleagues (2000) applied the consider the opposite technique to a real-life scenario during which participants were asked to provide a price estimate for a secondhand car. Participants, who were selected to be experts in cars, were asked to determine if the asking price for a car was reasonable or not based on information indicative of its value (the make of the car, how many miles were on the car, etc.). Using a two-by-two design, half the participants were given a high anchor and half were given a low anchor (i.e., 5000 German Marks vs. 2800 German Marks); moreover, half were asked reasons why the anchor value might have been inappropriate (i.e., the consider the opposite intervention). Analyses revealed that the consider the opposite strategy mitigated the anchoring effect and allowed participants to make more accurate estimates of the car's market value.

Another study (Nagtegaal et al., 2020) found that the consider the opposite strategy could also be used to reduce anchoring biases in workplace settings. In the study, managers in the United Kingdom were asked to estimate an employee's performance, which was then compared to a more objective rating of the employee's performance using previous performance reports. Managers received either a high or low anchor of the employee's performance rating and half then received a consider the opposite intervention, during which they were asked to list two

reasons why the anchor was too high or too low. In the high anchor condition, those in the consider the opposite condition tended to make lower estimates (closer to the objective performance score) when compared to the control group; conversely, in the low anchor condition, those in the consider the opposite group tended to make higher estimates (closer to the objective performance score) when compared to the control group. Overall, results showed that considering the opposite allowed managers to take a more objective stance and produce more accurate decisions about employee performance.

Given the prior effectiveness of the consider the opposite intervention on other forms of judgment biases, investigating whether the intervention approach works to reduce social biases is warranted. Specifically, the present study will apply the consider the opposite intervention to a form of social bias: judgment biases that favour more over less physically attractive people.

To measure attractiveness-based discrimination, the study will use a version of the Judgment Bias Task (JBT; Axt et al., 2018): a decision-making task where participants review and evaluate applicants for a hypothetical academic honour society. This task has been used to assess physical attractiveness bias and has consistently been shown to produce attractiveness bias in judgment (Axt & Lai, 2019; Axt et al., 2023). The qualifications given to applicants vary to create one class of more qualified applicants and one class of less qualified applicants. Within each level of qualification, applicants are also evenly divided into being paired with faces that were previously rated to be higher or lower in physical attractiveness.

A consider the opposite intervention will be randomly assigned to participants before completing the JBT. Specifically, before making the accept or reject decisions on applicants, participants will complete a consider the opposite intervention that involves the evaluation of four practice applicants. For each practice applicant, participants in the consider the opposite

condition will be asked to list reasons why someone else may have come to a *different* conclusion when evaluating the applicant (i.e., if the participant would accept the applicant, they would need to list reasons why someone else would have rejected the applicant). To increase similarity across experimental conditions, another condition will ask participants to think of reasons why someone else would arrive at the *same* conclusion. Previous studies have successfully used similar implementations of the consider the opposite strategy (Mussweiler et al., 2000; Nagtegaal et al., 2020).

The design of the JBT allows for different investigations as to how considering the opposite might impact behavior on the task. Specifically, the JBT is commonly analyzed using signal detection theory (SDT; Green & Swets, 1966), an approach to decision-making that considers the amount of *noise* and *bias* in a series of judgments. Noise is measured by the degree to which participants can differentiate between more and less qualified applicants and calculated by sensitivity (d'). Bias is measured using response criterion, which refers to decision threshold. A lower criterion value in this context suggests greater leniency (i.e., applicants do not have to be as qualified to receive an "accept" response). Attractiveness-based favouritism then emerges when more and less physically attractive targets differ in response criterion, and JBT performance consistently shows lower response criterion (greater leniency) for more versus less physically attractive targets (Axt et al., 2018; Axt et al., 2021).

This SDT approach to decision-making is useful to disentangle whether an intervention reduces noise (increases overall accuracy on the task), reduces bias (reduces favouritism towards one group over another), or impacts both outcomes simultaneously. That is, some interventions may lessen discrimination by reducing noise (i.e., reducing the total amount of errors made) and others may lessen discrimination by reducing bias (i.e., not impacting the total number of errors

made in judgment, but at least making those errors more evenly distributed between social groups). Although empirically distinct, it is also possible for a single intervention to reduce both noise and bias, such as during an intervention in Axt and Lai's (2019) study involving both a time delay in responses on the JBT (reducing noise) and warnings of attractiveness bias (reducing bias).

Given the nature of the consider the opposite intervention, I hypothesize that the manipulation will reduce both noise and bias on the JBT. Like past studies have shown (Mussweiler et al., 2000; Nagtegaal et al., 2020), the consider the opposite intervention results in participants making more accurate decisions compared to their first impressions by challenging their initial judgment. As a result, the intervention could be expected to improve sensitivity. The consider the opposite intervention may also guide participants away from any initial favouring of physically attractive applicants by asking participants to reason objectively about why someone may make an opposite decision. As a result, manipulation should also lessen the effects of physical attractiveness on judgments.

Finally, I will include an exploratory analysis investigating whether the effects of the manipulation differ for White vs. non-White participants. This analysis is included to investigate whether the impact of Eurocentric beauty standards produces different effects for White versus non-White participants (Robinson-Moore, 2008; Chen et al., 2020), since all the images in the JBT were of White targets.

Methods

Participants

550 participants completed this study through the Project Implicit research pool or the McGill SONA human participant pool and ranged in age from 18 to 83 (M = 30.61, SD = 14.10).

Participants were excluded from analyses if they accepted more than 80% of applicants or rejected less than 20% of applicants during the JBT. Participants were also excluded if they accepted or rejected every single more or less attractive applicant (Axt et al., 2018).

We initially targeted a sample size of 506 eligible participants, but data exclusions and difficulty with recruitment left a final sample of 480 eligible participants. This sample provided 80% power to detect a between-subjects effect as small as $d = .26$. Demographic information was collected when participants first registered for the Project Implicit research pool or for the undergraduate SONA pool. Slightly different demographic measures were given to participants from each sample source. See Table 1 for the demographic characteristics of the sample, broken down by each recruitment source. Study design, measures, and analyses were pre-registered. https://osf.io/zdyps

Table 1

Demographic Characteristics

Characteristics (SONA – 108 participants)	*N*	%
Race		
American Indian/Alaska Native	0	0
East Asian	24	22.2
South Asian	10	9.3
Native Hawaiian or other Pacific Islander	0	0
Black or African American	1	0.9
White	47	43.5
Mixed Race	12	11.1
Hispanic	1	0.9
Other/Unknown	13	12.0
Gender Identity		
Female	80	74.1
Male	20	18.5
Trans female/Trans woman	1	0.9
Trans male/Trans man	0	0
Genderqueer/Gender nonconforming	3	2.8
Other/Unknown	4	3.7
Characteristics (PI – 372 participants)	*N*	%
Race		
Asian	27	7.3
Black	40	10.8
Hispanic	43	11.6
Middle Eastern	3	0.8
Mixed Race	18	4.8
White	240	64.5
Other/Unknown	1	0.3
Gender Identity		
Female	238	64.0
Male	117	31.5
Other	9	2.4
Unknown	8	2.2

Procedure

The study followed an experimental design with two conditions: participants were randomly assigned to a "consider" or a "consider the opposite" intervention.

The study used the Judgment Bias Task (JBT; Axt et al., 2018) as the primary outcome measure. In this version of the JBT, participants accepted or rejected 64 applicants for an honours committee based on provided profiles. Using a two-by-two design, applicants on the JBT differed on two dimensions: 1) more or less academically qualified and 2) more or less physically attractive.

The JBT contained three phases: an initial screening phase where participants saw all the applicants in quick succession, an intervention phase where participants received either the "consider" or "consider the opposite" intervention, and the test phase where participants chose to accept or reject each applicant based on the picture and information given. The intervention phase included the intervention itself and four sample trials prompting the participants through the steps necessary to complete the intervention. See Table 2 for the intervention text. The four sample trials were novel profiles that did not appear in the testing phase, and the four sample trials were shown without an accompanying face. After completing the JBT, participants were asked to fill in questionnaires about their perceived and desired task performance and their explicit attractiveness attitudes (Axt et al., 2018).

When all aspects of the study were completed, participants were given an overview of their performance on the JBT. Specifically, this feedback detailed the number of applicants they had accepted or rejected broken down by high vs. low levels of qualifications and high vs. low levels of physical attractiveness.

Table 2

Intervention Text Presented in the JBT

Intervention	Intervention Text
Consider	"Please try your best to select the most qualified applicants. As you select the applicants, please continue to think about why someone else could have made the SAME decision as you."
Consider the Opposite	"Please try your best to select the most qualified applicants. As you select the applicants, please continue to think about why someone else could have made the OPPOSITE decision as you."

Measures

Judgment Bias Task (JBT)

The Judgment Bias Task (JBT; Axt et al., 2018) was used to assess biased judgment towards more versus less physically attractive people. Before starting the task, participants viewed 64 profiles of applicants for an honours committee one after another. Profiles of applicants varied in two levels of qualifications and two levels of attractiveness. Qualification level, which was measured using four different criteria, divided participants into less qualified applications and more qualified applicants. These four criteria (i.e., science GPA, humanities GPA, references, and interview scores) were each assigned a point value on a scale from one to four. GPA was directly translated onto this point scale (e.g., GPA of 3.2 would be equal to 3.2 points), recommendation letters were scored from poor to excellent (poor = 1, fair = 2, good = 3, excellent = 4), and interview scores ranging from 0 to 100 were calculated by dividing the total score by 25 (e.g., an interview score of 90 would equal 90/25 = 3.6). Less qualified profiles had a total score of 13 and more qualified profiles of 14.

Profiles also differed on two levels of physical attractiveness, which was displayed using a profile picture. Attractiveness was tested using a sample of undergraduate students ($N = 63$) such that attractive profiles were rated more highly than unattractive profiles using a between-subjects comparison, and all attractive profiles were determined to be significantly higher rated

than unattractive profiles ($d > 2.00$; Axt et al., 2018). To remove any influence of applicant race, all stimuli on the JBT were of White people.

During an initial screening phase, all 64 applicants' profiles were shown for one second each. This screening phase is included so that participants can see the range of qualifications in the applicant pool before making any accept or reject decisions. The "consider" and "consider the opposite" interventions occurred immediately after the screening phase. After the interventions, participants entered the selection phase, where they chose whether to accept or reject each applicant based on the picture and information given. Each participant was randomly assigned to one of 12 JBT orders; across orders, each face was equally likely to be paired with a more or less qualified profile, and each profile was equally likely to be paired with a more or less physically attractive face.

Perceived and Desired JBT Performance

Perception of performance was measured using a self-report question. Participants were asked to rate their perceived bias towards physically attractive/unattractive applicants on the JBT using a 7-point scale from "I was extremely easier on physically unattractive applicants and extremely tougher on physically attractive applicants" (−3) to "I was extremely easier on physically attractive applicants and extremely tougher on physically unattractive applicants" (+3). Desired JBT performance was measured using a 7-point scale from "I wanted to be extremely easier on physically unattractive applicants and extremely tougher on physically attractive applicants" (−3) to "I wanted to be extremely easier on physically attractive applicants and extremely tougher on physically unattractive applicants" (+3).

Explicit Attractiveness Attitudes

Explicit attractiveness attitudes (Axt, 2017) were measured with a single item that used a 7-point scale ranging from “I strongly prefer physically unattractive to physically attractive people” (−3) to “I strongly prefer physically attractive to physically unattractive people” (+3).

Results

Within each condition, we first ran one-sample *t*-tests to examine whether sensitivity scores were reliably higher than zero (i.e., above chance). Both conditions showed evidence of above-chance responding. For the consider condition, the one-sample *t*-test revealed that the mean (M = .70, SD = .15) was significantly higher than chance responding ($t(245) = 21.18$, p = <.001). For the consider the opposite condition, the one-sample *t*-test also revealed that the mean (M = .72 SD = .15) was significantly higher than chance responding ($t(233) = 21.56$, p = <.001). A two-tailed independent-samples *t*-test found no reliable difference in sensitivity between the between the consider and consider the opposite conditions ($t(478) = -.90$, p = .371, d = -.08), indicating that sensitivity was not reliably improved or worsened by the consider the opposite intervention.

We next ran a two-tailed within-subjects *t*-test to investigate whether criterion differed between more and less physically attractive applicants (i.e., whether there is an attractiveness bias in criterion) within each condition. The within-subjects *t*-test for the consider condition compared the mean scores of the attractive group criterion (M = -.18, SD = .49) and the mean scores of the less attractive group criterion (M = -.05, SD = .54). The analysis revealed a significant difference between means, $t(245) = -4.54$, p = <.001, d = -.29, indicating that there was in fact bias that gave more leniency towards physically attractive applicants. The within-subjects *t*-test for the consider the opposite condition also compared the mean scores of the attractive group criterion (M = -.18, SD = .49) and the mean scores of the less attractive group

criterion (M = -.12, SD = .50). The analysis also revealed a significant difference between means (t(233) = -2.14, p = .034, d = -.14), indicating that there was again a bias that favoured the more attractive (but equally qualified) applicants.

Next, a comparison of levels of criterion bias was completed using a two-tailed independent-samples t-test between conditions. In this analysis, a criterion bias difference score between more attractive applicants and less attractive applicants was calculated, such that higher values meant greater leniency for more relative to less physically attractive applicants. The independent-samples t-test indicated a non-significant difference between the consider (M = .13, SD = .45) and consider the opposite (M = .06, SD = .41) conditions (t(478) = 1.82, p = .069, d = .17), signifying that the consider the opposite intervention did not reliably reduce attractiveness bias, though this effect could be considered marginally significant.

Finally, an exploratory analysis of race effects on physical attractiveness bias was conducted, which explored whether non-White (n = 181) participants displayed more or less bias when compared to White (n = 287) participants in both conditions. This analysis was done using an independent-samples t-test examining the difference between non-White and White participants' criterion bias. Note that mixed-race White participants (n = 12) were excluded from this analysis.

First, a comparison of levels of criterion bias for White participants was done using a two-tailed independent-samples t-test between conditions. The independent-samples t-test indicated a significant difference between the consider (M = .17, SD = .49) and consider the opposite (M = .01, SD = .38) conditions (t(285) = 3.13, p = .002, d = .37), signifying that the consider the opposite intervention reliably reduced attractiveness bias among White participants. Among non-White participants, the independent-samples t-test did not indicate a significant

difference between the consider (M = .06, SD = .35) and consider the opposite (M = .12, SD = .45) conditions ($t(179) = -0.99$, $p = .323$, $d = -.15$), signifying that the consider the opposite intervention did not reliably reduce attractiveness bias among non-White participants.

Discussion

The purpose of this study was to investigate whether a consider the opposite strategy could be used to reduce physical attractiveness bias, as measured by performance on the Judgment Bias Task. Using a signal detection approach, I expected the consider the opposite strategy to reliably reduce physical attractiveness bias, represented by a reduced criterion bias to accept physically attractive applicants compared to less physically attractive applicants. I also expected a decrease in noise in the consider the opposite condition. Following this, an exploratory analysis was done to investigate whether there were any significant differences between White and non-White participants in physical attractiveness bias and intervention effectiveness.

Analyses found that the consider the opposite approach did not reliably reduce criterion bias, at least when tested among the full sample. Although this finding was not in line with initial hypotheses, it does not directly refute any existing literature due to the novelty of the consider the opposite approach being applied to the domain of attractiveness-based biases. However, it is perhaps surprising that the intervention failed to impact criterion bias or sensitivity given the intervention's success at reducing other types of judgment biases (e.g., Arkes et al., 1988; Mussweiler et al., 2000).

Understanding why the intervention failed to produce consistent effects on judgment may come from a deeper exploration into why the intervention has been effective in prior work. There are currently three working theories as to how the "consider the opposite" strategy works for

reducing judgment biases. The first potential mechanism suggests that it reduces anchoring effects (i.e., a relatively inflexible reference point around which an individual makes decisions; Kahneman & Tversky, 1974). In contexts where judgments are made with uncertainty, all judgments tend to fall close to an initial anchoring point (Kahneman & Tversky, 1974). Considering opposite scenarios may then move judgments further from this anchoring point and can therefore increase judgment accuracy in many contexts (Lord et al., 1984). The second potential mechanism proposes that considering more and opposite outcomes creates uncertainty in one's decisions, which tends to lead to more neutral (and unbiased) stances being taken (Hirt & Markman, 1995). Finally, the third potential mechanism theorizes that decisions are made based on the ease of outcome generation (Kahneman & Tversky, 1982). In other words, we think that something is more likely to occur or be correct based on how easily we think of it. In this case, considering outcomes generated with more difficulty, which follows from considering the opposite course of action, will decrease the likelihood of jumping to conclusions based on our first impressions.

Considering these competing explanations, it is possible that the mechanism used by the consider the opposite intervention is not one that effectively reduces bias in intergroup contexts, or at least in attractiveness-based biases. In that case, biases caused by physical attractiveness, or the halo effect, may simply function too differently from other forms of judgment biases and require different intervention approaches. It has been noted in past studies that the consider the opposite technique has produced mixed results for reducing confirmation biases (Whitt et al., 2023; Lilienfeld et al., 2009), which may also function similarly to the halo effect. While some studies have shown that perspective-taking techniques have decreased confirmation bias (Lilienfeld et al., 2009), others have found the techniques to be ineffective. For example, Whitt

and colleagues (2023) tested three cognitive debiasing techniques for reducing the halo effect: consider the opposite, psychoeducation about bias, and social norm manipulation. Across five different topics, only the social norm manipulation was shown to be somewhat effective at reducing confirmation biases, supporting the ineffectiveness of the consider the opposite intervention. Attractiveness-based biases on the JBT may then function similarly to a confirmation bias (Axt & Johnson, 2021), where participants follow an initial inclination to favor more over less physically attractive applicants, and as a result may be resistant to "consider the opposite" interventions. Further research is warranted to investigate whether this debiasing technique could function differently with other intergroup biases, such as racial bias or gender bias which may have more clearly defined in-group and out-groups.

In addition to differences between judgment biases and intergroup biases, there is also the possibility that the current sample was not sufficient to produce a reliable effect. For one, the sample used in the current study was slightly smaller than the target sample. At only 480 eligible participants, the current sample misses the 506 participants required to reach the targeted statistical power of 80% for a between-subjects effect size of $d = .25$. Lower statistical power influences the reliability of the results, and a larger sample size may reveal results not able to be uncovered within the current sample. Second, our sample may have been inappropriate given their lack of expertise in the judgment domain being investigated. Previous studies showing effective uses of the consider the opposite intervention have relied on expert samples, such as car experts (Mussweiler et al., 2000) and hiring managers (Nagtegaal et al., 2020). Although several studies have similarly used laypeople (Greitemeyer, 2023; Chandon & Wansink, 2007), it is possible that experts may be more inclined to consider the objective academic information presented in the JBT compared to laypeople and may be more responsive or engaged with the

instructions to consider opposite perspectives. Investigating whether stronger results emerge among expert samples (e.g., professional admissions officers) should be a focus for future research on this topic, as it is also more relevant to real-world applications.

Since results indicate that the consider the opposite intervention did not produce strong evidence of reducing physical attractiveness bias in academic selection scenarios, hiring or acceptance committees should be wary of incorrectly believing that this strategy reduces bias. Instead, managers may want to continue to rely on discrimination-reduction strategies previously shown to work, such as time delays and implementation intention strategies (Axt et al., 2023). However, the marginally significant results from the criterion bias analysis within the full sample suggest that this research would greatly benefit from a high-powered direct replication that could produce more conclusive data about the effectiveness of the intervention.

At the same time, the sensitivity analyses also did not reveal any reliable differences between the consider and consider the opposite conditions. Initially it was suspected that producing judgment-incongruent reasons (i.e., considering opposite reasons from the initial judgments would take longer) would cause a significantly longer time delay in decision-making, increasing sensitivity in the consider the opposite condition. However, the lack of a reliable effect on sensitivity suggests that participants in the consider the opposite intervention may have experienced similar benefits as those in the "consider" condition. Including a pure control condition in future research would help identify whether simply being asked to further deliberate on one's judgment (either to arrive at the same *or* a different conclusion) increases judgment accuracy (e.g., Axt & Lai, 2019).

Interestingly, our exploratory race-based analysis yielded mixed results. In particular, the consider the opposite intervention worked to significantly reduce criterion bias among White

participants but not non-White participants. Since previous studies have shown that non-White participants also show physical attractiveness bias on the JBT (Axt et al., 2018), this effect could be attributed to the consider the opposite intervention. The difference between White and non-White participants may come from the interaction between the outgroup homogeneity effect (Quattrone & Jones, 1980) and ease of counter-stereotype generation (Kalawami et al., 2000; Power et al., 1996). As the outgroup homogeneity effect causes individuals to think of the outgroup (in this case, people of a different race) as less diverse, it may be more difficult for them to think of counter-stereotypes or opposite explanations (Quattrone & Jones, 1980). This may be a factor for why the consider the opposite condition was only effective among White participants (as the applicants were also White and therefore considered the in-group). Note that by splitting these analyses into smaller categories, sample sizes and statistical power are much lower.

This study is not without limitations. As mentioned before, the sample size was insufficient to reach the desired statistical power. In addition, the sample was made up of mostly women, which could again underestimate the relationship between attractiveness and bias, as previous studies have shown that while physical attractiveness bias exists for both men and women, it presents more strongly in men (Murphy et al., 2015). A weaker initial attractiveness bias, due to the majority female sample, may have diminished the effects of the consider the opposite intervention.

Finally, due to the study's nature, participant compliance with the methodology was less rigorous than desirable. The study was done online, independently, and without supervision. Under these conditions, it is hard to ensure that the participants implemented the consider the opposite or consider interventions for each participant. Given the limited time frame (20

seconds) that participants had to make decisions on the JBT, it is possible that participants reverted to instincts or first impressions after the first few applicants. To solve this problem, future work could increase the amount of time that participants had to think about each applicant (increasing the amount of time available to fully consider or consider the opposite) and ask participants to continue to report their consider/consider the opposite reasons in writing, as in the practice trials. However, there are downsides to these more immersive manipulations, mainly in that they increase the amount of time required to complete the study. Such a change could then reduce the number of participants willing to complete the task and consequently, our sample size would have been much smaller and even less representative.

Another solution would be to do the study in person or require the participants to verbally reason why they would consider or consider the opposite after each participant, as done in Mussweiler and colleague's study on car experts (2000). This alternative approach would increase enforcement of the intervention but would require the study to be either in person (decreasing the possible diversity within the sample) or would require additional video or audio recording permissions. Regardless, increasing compliance with the interventions would have likely increased the intervention's strength on physical attractiveness bias, or at minimum would have strengthened our confidence in results.

Building on the limitations, future studies should replicate this study on a larger scale to reach greater statistical power and follow up on the marginally significant effect reported here on criterion bias. In addition, this study could be replicated with an expert population (participants who work on acceptance or hiring committees) to better reflect how this intervention would work in real-life situations. It is likely that laypeople would never need to make acceptance or hiring decisions in their day-to-day lives, making the results of this study less relevant when compared

to experts. Finally, future research can build upon the exploratory analysis examining race-based effects by using a racially diverse sample of applicants on the JBT. This would better mimic real-life scenarios and give researchers a better idea of how attractiveness bias presents in day-to-day hiring practices.

The current study contributes to the existing debiasing literature by applying techniques from judgment bias fields of research to an area of intergroup bias research. This work builds on novel efforts to integrate debiasing techniques proven effective in one area of bias research to other areas of bias research to increase the number of available debiasing techniques (e.g., He et al., 2021). Results found that the consider the opposite intervention did not reliably reduce attractiveness-based biases in judgment. This may indicate that certain intergroup biases, such as attractiveness bias, function too differently from judgment biases and therefore cannot employ similar debiasing techniques, though additional tests of this approach in larger samples or different operationalizations would strengthen this conclusion. Examining the collection of techniques that can and cannot be carried over between fields may give researchers a better understanding of mechanistic differences between judgment and intergroup biases. Future efforts in this area should continue to investigate which techniques can be carried over between fields of bias research.

References

Agthe, M., Strobel, M., Spörrle, M., Pfundmair, M., & Maner, J. K. (2016). On the borders of harmful and helpful beauty biases. *Evolutionary Psychology*, *14*(2), 147470491665396. https://doi.org/10.1177/1474704916653968

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. https://doi.org/10.1037/0021-9010.73.2.305

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

Axt, J. R., & Johnson, D. J. (2021). Understanding mechanisms behind discrimination using diffusion decision modeling. *Journal of Experimental Social Psychology*, *95,* 104134. https://doi.org/10.1016/j.jesp.2021.104134

Axt, J. R., Yang, J., & Deshpande, H. (2022). Misplaced intuitions in interventions to reduce attractiveness-based discrimination. *Personality and Social Psychology Bulletin*, *49*(4), 527–540. https://doi.org/10.1177/01461672221074748

Barrick, M. R. & Mount, M. K. (2012). Select on Conscientiousness and Emotional Stability. In Handbook of Principles of Organizational Behavior, E.A. Locke (Ed.). https://doi.org/10.1002/9781119206422.ch2

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of behavioral economics* -

Foundations and applications 2 (pp. 69–186). North-Holland/Elsevier. https://doi.org/10.1016/bs.hesbe.2018.11.002

Chen, T., Lian, K., Lorenzana, D.J., Shahzad, N., & Wong, R. (2020). Occidentalisation of Beauty Standards: Eurocentrism in Asia. https://doi.org/10.5281/zenodo.4325856

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Fischhoff B. (1982). Debiasing. In Kahneman D., Slovic P., Tversky A. (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.

Green, D. W., & Swets, J. A. (1966). *Signal detection theory and Psychophysics*. John Wiley & Sons.

He, J. C., Kang, S. K., & Lacetera, N. (2021). Opt-out choice framing attenuates gender differences in the decision to compete in the laboratory and in the field. *Proceedings of the National Academy of Sciences*, *118*(42). https://doi.org/10.1073/pnas.2108337118

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*(6), 1069–1086. https://doi.org/10.1037/0022-3514.69.6.1069

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

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. *Judgment under Uncertainty*, (pp. 201–208). https://doi.org/10.1017/cbo9780511809477.015

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*(5), 871–888. https://doi.org/10.1037//0022-3514.78.5.871

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*(4), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x

Lord, C. G., Lepper, M. R., & Preston, E. (1984a). Considering the opposite: A corrective strategy for Social Judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Lord, C. G., Lepper, M. R., & Preston, E. (1984b). Considering the opposite: A corrective strategy for Social Judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Maestripieri, D., Henry, A., & Nickels, N. (2016). Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/s0140525x16000340

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, *81*(1), 11–21. https://doi.org/10.1037/0021-9010.81.1.11

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*(4), 512–523. https://doi.org/10.1177/0146167210362789

Murphy, C., Hussey, T., Barnes-Holmes, D., & Kelly, M. E. (2015). The implicit relational assessment procedure (IRAP) and attractiveness bias. *Journal of Contextual Behavioral Science, 4*(4), 292–299. https://doi.org/10.1016/j.jcbs.2015.08.001

Power, J. G., Murphy, S. T., & Coover, G. (1996). Priming prejudice how stereotypes and counter-stereotypes influence attribution of responsibility and credibility among ingroups and outgroups. *Human Communication Research*, *23*(1), 36–58. https://doi.org/10.1111/j.1468-2958.1996.tb00386.x

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology*, *38*(1), 141–152. https://doi.org/10.1037//0022-3514.38.1.141

Robinson-Moore, C.L. (2008). Beauty Standards Reflect Eurocentric Paradigms-So What? Skin Color, Identity, and Black Female Beauty.

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

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

Tajfel, H., & Turner, J.C. (1979). An integrative theory of intergroup conflict.

Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PLOS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0148284

Whitt, C. M., Tullett, A. M., Bowes, S. M., McDiarmid, A. D., Lilienfeld, S. O., & Hart, W. (2023). Weakening the ideological immune system: Can debiasing techniques reduce confirmation bias? *Comprehensive Results in Social Psychology*, *7*(1–2), 1–24. https://doi.org/10.1080/23743603.2023.2196755

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142. https://doi.org/10.1037/0033-2909.116.1.117

Statement of Contribution

AG conducted the literature review, analyzed the data, and interpreted the results. JA and AG contributed to designing the study and developing the procedures and stimuli. JA collected the data and entered and coded the data.