

Exploring the Longitudinal Effects of Discrimination Interventions

Anya Williamson

Department of Psychology, McGill University

PSYC 380D: Honors Psychology Research Project

Under the supervision of Eliane Roy & Dr. Jordan Axt

April 25, 2024

Abstract

Many studies have demonstrated the usefulness of interventions to reduce discrimination in a variety of domains. However, the long-term effects of these interventions have often show less consistent results. In this study, we tested two previously successful discrimination interventions longitudinally on a decision-making task that has demonstrated a consistent attractiveness-based bias. Participants completed the decision task pre-intervention and two to four days post-intervention. We found that both interventions, Two-Out-of-Three Rule and Criteria Reinforcement Exercise, significantly increased sensitivity and reduced criterion bias in participants over two days after the initial treatment. These findings emphasize the importance of simple criteria as a means of reducing discrimination over time. Furthermore, the flexibility of these interventions indicates a strong potential for real-world applications where decision-making tasks that have objectively correct and incorrect answers are influenced by biased and socially irrelevant information.

Exploring the Longitudinal Effects of Discrimination Interventions

While recent research on implicit attitudes has identified a general decrease in implicit bias in many domains, such as sexuality or race, (Charlesworth & Banaji, 2022), these trends have not always been accompanied with behavioural changes. Indeed, discrimination, which refers to the unfair treatment of individuals based on their traits, characteristics, or group membership (Monk et al., 2021), is a problem that remains rampant. Discrimination can take many forms, such as unfair employment outcomes (Lippens et al., 2023; Nault et al., 2020; Monk et al., 2021), biased health treatment (Williams et al., 2019), or lesser housing opportunities (Fibbi et al., 2021). As such, this pervasive and systemic problem impacts a number of social groups in a wide variety of contexts. At the individual level, discriminatory behavior is often correlated with explicitly held biases and stereotypes, which represent the consciously held cognitive and affective mental representations one has of others (Lai et al., 2013; Fibbi et al., 2021). However, discriminatory behaviour can also be influenced by ones' implicit attitudes, making behavior change a more difficult task (Wallaert et al., 2010). Altogether, discrimination is a problematic yet multifaceted issue that needs to be better understood to reduce inequalities.

One very influential, yet often understudied, form of discrimination is based on attractiveness favouritism, wherein more physically attractive people tend to receive more favourable treatment relative to less physically attractive people (Lippens et al., 2023). Attractiveness-based biases are present in a number of social fields, from education (Talamas et al., 2016), to social capital (Gordon et al., 2013), to employment (Scholz & Sicinski, 2015; Mobius & Rosenblat, 2006). In the popular media, many have heard of the *Halo* effects (Thorndike, 1920) which describes the rather positive expectations and treatment afforded to

more attractive people over others. However, less attention has been paid to the fact that discrimination is relative: the favorable treatment of attractive individuals comes at the unfair treatment of individuals considered less attractive (Fibbi et al., 2012). Researchers found that more attractive individuals were more likely to earn higher wages and to be recommended for promotions at work (Nault et al., 2020). They were also more likely to be elected as a parliamentary candidate (Berggren et al., 2010), or to receive higher quality and easiness ratings on student evaluations of professors (Felton et al., 2008). However, no links have been found between attractiveness and mental ability (Feingold, 1992), meaning that the benefits that individuals receive for being considered more attractive do not hinge on greater competency or merit relative to those who are considered less attractive.

While a large body of discrimination research has been devoted to studying racially motivated prejudice (Lai et al., 2016; Gaddis, 2015), or gender-based discrimination (Adamovic & Leibbrandt, 2022), comparatively less research has been done on attractiveness-based discrimination. One recent meta-analysis found that physical attractiveness was almost equally as impactful a category towards treatment in hiring as having a disability, being an older applicant, or belonging to an ethnic minority (Lippens et al., 2023). While racial, ethnic, and national minorities received nearly one third less positive responses to employment applications than majority counterparts, physically less attractive applicants resulted in an average of two-fifths lower positive responses (Lippens et al., 2023). In other words, physically less attractive individuals are often unfairly disadvantaged over more attractive individuals.

One element that can make reducing attractiveness-based discrimination particularly difficult is its relatively subconscious nature as compared to other forms of discrimination (Axt et al., 2018; Lai et al., 2013). For example, self-reported attitudes towards black people tend

to be moderately associated with measures of implicit attitudes, such as those measured by the Implicit Association Test (Greenwald et al., 1998; Lai et al., 2013). On the other hand, explicit attitudes regarding physical attractiveness tend to be less correlated with behavioral outcomes (Axt et al., 2018). This means that, despite an often-outspoken desire to treat more and less physically attractive individuals equally on a decision-making task, people are still influenced to behave in discriminatory ways (Axt et al., 2018). However, even implicitly held individual beliefs, although they tend to be more stable, can be malleable to lab-based interventions (Lai et al., 2013). A number of prejudice-reduction mechanisms have been identified, with some interventions finding more success than others. A meta-analysis on implicit prejudice interventions found that certain strategies such as retraining underlying associations had mixed success, while shifting the context of evaluation and controlling behavioral strategies could only reduce prejudice scores the day-of (Lai et al., 2013). However, several meta-analyses about prejudice reduction have also indicated the possibility that publication biases could amplify the effects of these interventions (Paluck et al., 2021; Forscher et al., 2019).

As a consequence of these mixed results, some interventions have started to focus only on shaping the discriminatory behavior itself rather than the attitudes held by individuals. Some of these more behavioral interventions follow the lines of implementation intention strategies that provide an “if [blank], then [blank]” script for participants to follow, explicitly detailing the exact actions one must take if a certain situation or context occurs (Wallaert et al., 2010; Lai et al., 2016). Another behavioral strategy that has seen success is by limiting or blinding the socially irrelevant information, such as race, gender, socioeconomic status, or attractiveness, during decision-making tasks so that participants are less biased (Axt & Lai, 2019). Giving participants clear instructions provides a heuristic framework that they can rely on during

decision-making tasks rather than underlying prejudices that may influence them.

While the findings from empirical intervention testing are important for advancing the field of discrimination reduction, their long-term effects are severely understudied. In a meta-analysis of experiments to change implicit associations (Forscher et al., 2016), only 22 of the 585 studies analyzed the longitudinal effects of interventions and, of those 22, only 9 found significant evidence of lasting change. Another study by Lai et al. (2016), found that none of nine interventions that were previously successful at reducing racial implicit prejudice the day-of held statistically significant effects over time. These findings show that a key issue in the study of longitudinal effectiveness of interventions is the difference between malleability and change. While malleability refers to shifts in evaluation that are limited to the immediate situational context, change refers to shifts in evaluation that persist across multiple situational contexts (Nosek et al., 2012). Indeed, while individual attitudes are malleable in experimental settings, they are often interpreted as change despite long-term change being harder to demonstrate (Han et al., 2010). As such, the longitudinal effectiveness of discrimination interventions remains a gap in the literature that requires further exploration.

The Present Work

In this thesis, we investigated the longitudinal effectiveness of two previously successful discrimination interventions on a decision-making task that has consistently shown an attractiveness-based bias, the Judgment Bias Task (JBT; Axt et al., 2018). The version of the JBT used in this study asked participants to evaluate applicants for an academic honor society. During the task, participants are presented with vignettes that include a picture of the applicants, and four levels of qualification: their science GPA (scale of 1-4), their humanities GPA (scale of 1-4), their interview score (scale of 1-100), and their recommendation letter quality (poor, fair,

good, or excellent). Each face on the JBT represents either more or a less physically attractive applicant that have been pre-tested and validated in previous work (Axt et al., 2018).

First, we present a pilot study which tested whether the participants exhibited a practice effect when completing the JBT twice in the same study session. This pilot study aimed to improve the validity of data collected during the latter stage of the project by quantifying the potential impact of practice effects as a confounding variable (i.e., whether any improvements on JBT performance following an intervention could be attributed to practice effects emerging from simply completing the measure twice). We hypothesized that there would not be a practice effect between participants' first and second JBT performances within the same sitting.

Next, we tested two interventions (Two-out-of-three-rule, Criteria Reinforcement exercise; Roy et al., under review) that have been shown to effectively reduce discrimination based on physical attractiveness or political affiliation when administered immediately before the JBT. We hypothesized that these interventions would reduce discrimination from the participants' first JBT attempt relative to their second JBT attempt a few days later.

Pilot study

Methods

Participants

We recruited 196 undergraduate participants from the McGill University Participant Pool and compensated them with course credit. As in prior work (Axt et al., 2018; Axt & Lai, 2019), participants were excluded from the analyses if they accept less than 20% or more than 80% of all applicants on the JBT. They were also excluded if they accepted or rejected every more or less physically attractive applicant. Lastly, participants were removed from analyses if they failed the attention check item (a single item at the end of the study that asked participants if they

paid sufficient attention to be included in analyses). The final sample was then $N = 178$, and provided us with 80% power to detect a within-subjects effect size of $d = .27$. All data exclusions and analyses were pre-registered at (https://osf.io/f27gp/?view_only=464f033fda734b0dbd91f5c494770401). Complete data, analysis script, and materials are all available at (https://osf.io/n75j6/?view_only=57f44b54678a44529728d5193837c425).

Measures

Judgement Bias Task. In the JBT, 64 applicants are arranged in a two-by-two design; half of the applicants are paired with a picture of either more or less physically attractive person, and half are categorized either more or less qualified. The sum of the applicants' qualifications is equal amongst their qualification category (more qualified versus less). More specifically, when the information for each qualification is set to a scale of 1-4, the sums of each applicant's qualifications add up to 13 for less qualified applicants and 14 for more qualified applicants. See Axt et al. (2018) for more details on the paradigm. At the beginning of the task, participants are instructed to accept approximately half of applicants into the academic honors society and to weigh all qualification indicators equally. Participants are then presented with a 60-second encoding phase where they are shown each applicant for one second at a time to gauge the overall range of qualifications. During the decision phase of the JBT, participants are given up to 15 seconds to make a 'reject' or 'accept' decision for each of the 64 applicant trials.

JBT performance is analyzed using Signal Detection Theory (Correll et al, 2002; Green & Swets, 1966; Axt & Lai, 2019). First, sensitivity (d') represents the ability for participants to accurately accept the more over the less qualified applicants. A higher sensitivity score indicates that participants successfully accepted more qualified applicants and rejected less qualified

applicants. Next, criterion represents the decision threshold for accepting applicants, with lower values meaning more leniency. We calculate a criterion bias score, which is a difference score between criterion for the more versus less qualified candidates. As such, a higher criterion bias indicates that participants are more likely to accept more attractive applicants over less attractive applicants when they have the same qualifications. For this pilot study, we will compare participant's JBT outcomes (sensitivity and criterion bias) between their first and second performance to determine whether practice changes performance on the JBT.

Self-Report Questionnaire. A three-item self-report questionnaire was administered at the end of the study. This questionnaire included one item about how the participants thought they performed on the task (i.e., whether they believe they showed favoritism towards more versus less attractive candidates). Another item asked how they wanted to perform on the task (i.e., whether they were trying to favor one group of candidates over another). Lastly, participants were asked whether they prefer more or less physically attractive people on a scale from 1-7, with 4 being the neutral point. Before completing the debriefing, participants completed a 7-item demographics questionnaire. The self-reported items and demographics were not included in any primary analyses but were added to the study to facilitate potential exploratory analyses in future work (e.g., whether interventions that reduced discrimination were mediated by changes to desired JBT performance).

Procedures

All participants completed the 64-trial JBT twice in the same study session. Once they completed the first JBT, participants repeated the same randomized order of the task. Lastly, participants completed the self-report and demographics questionnaires. All participants were debriefed at the end of the study.

Results

In the Pilot Study, we examined differences in sensitivity and criterion bias from participants' first JBT attempt to their second to establish whether JBT results would be influenced by practice effects. First, we evaluated whether sensitivity was significantly different than 0 in both JBT sessions, and found significant differences (first session: $t(177) = 27.987$, $p < 0.001$, $d = 2.098$, 95% $CI = [0.95; 1.09]$; second session: $t(177) = 29.893$, $p < 0.001$, $d = 2.241$, 95% $CI = [0.86; 0.98]$). Then, we conducted a paired t-test comparing sensitivity in the first versus the second, and found significant difference between the two, $t(177) = 2.8947$, $p = 0.004$, $d = 0.23$, 95% $CI = [0.03, 0.17]$. Here, sensitivity significantly *decreased* from the first JBT session ($M = 1.02$, $SD = .49$) to the second ($M = 0.92$, $SD = .41$), although this effect was modest in strength.

Next, we compared criterion values for more versus less physically attractive applicants to determine if there was a significant criterion bias in each JBT attempts. Using within-subjects t-tests, we found that participants were more lenient towards more physically attractive applicant in both their first ($t(177) = 4.951$, $p < 0.001$, $d = .35$) and second ($t(177) = 5.904$, $p < 0.001$, $d = .41$) JBT attempts. Then, we used a paired t-test to explore whether criterion bias scores were significantly different between the first and session JBT attempts. We found no significant difference, $t(177) = 0.56$, $p = 0.58$, $d = 0.05$, 95% $CI = [-0.05; 0.08]$ in criterion bias between the first and session JBT. In other words, participants did show an attractiveness bias in both JBT attempts, and its magnitude did not reliably change over time.

Study 1

The Pilot Study indicated that sensitivity decreased from the participants' first JBT to their second, whereas there was no significant difference in criterion bias between the

participants' first and second JBTs. As the JBT attempts conducted one right after the other, the decrease in sensitivity may be attributed to a fatigue effect, as participants became less focused on accurately parsing through applicants. These findings also indicate that there is no presence of a "practice effect" for the JBT, meaning we could conduct the longitudinal study without concerns for practice confounding our results. For Study 1, we investigated the longitudinal effects of previously successful discrimination interventions (Roy et al., under review) to expand on their mechanisms of action and gain further insight into their potential long-term applications.

Methods

Participants

We recruited a total of 279 undergraduate participants from two university participant pools ($n_1 = 223$, $n_2 = 56$) and compensated them with course credit for completing each part of the study. We combined these two samples for analysis, since the study designs were identical, and more participants were needed to achieve enough statistical power in our analyses.

Participants were excluded from the data set following the same eligibility criteria as in the pilot study, as well as if they did not complete the second part of the study. These exclusion procedures provided us with a final sample of 222 participants which granted an 80% power to detect a small within-subject effect size of $\eta^2 = 0.015$. Participants were randomly assigned to one of two intervention conditions: *Two-Out-of-Three-Rule* or *Criteria Reinforcement Exercise*. All data exclusions and analyses were pre-registered. The pre-registration, complete data, analysis script, and materials are all available at

(https://osf.io/f27gp/?view_only=57f44b54678a44529728d5193837c425).

Procedures

Participants first completed the same 64-trial JBT used in the previous study, then were randomly assigned to one of two intervention conditions. During these interventions, participants were asked to remember a certain set of instructions for their second JBT attempt. In the *Two-Out-of-Three Rule* intervention, participants were informed that decisions are often biased by irrelevant characteristics that are visible from a person's photo. Participants were instructed to apply a simple decision rule to avoid this bias. Specifically, participants were told that applicants should be admitted if they fulfilled at least two of three criteria: a science GPA of 3.4 or better, an excellent recommendation letter, or an interview score of 81 or better. Participants completed three practice rounds where they received feedback on whether they applied the rule successfully. All faces and qualification values presented were novel and not used in the JBT. In the *Criteria Reinforcement Exercise* intervention, participants are provided with the averages of the candidate pool's science GPAs (3.47), humanity GPAs (3.44), recommendation letter ratings (between good and excellent), and interview scores (78.1). Participants were told to accept the candidates who were above average overall when using these four components. The intervention had an eight-trial practice round with novel applications that provided feedback on each trial.

Forty-eight hours after completing part one of the study, participants received a link to complete the second part of the study within the next two days. Before starting the second JBT, participants were reminded that they received an intervention at the end of the first part of the study. However, they were not given exact prompts again, and were instead instructed to "As best as you can, please remember what information was given to you in that activity and apply it when completing the upcoming decision-making task". The second JBT followed the same 64-

trial structure as previously described. Finally, participants completed the self-report items and demographics questionnaire before they were fully debriefed about the purpose of the study.

Results

We conducted two two-factor mixed ANOVAs to determine the effects of both intervention conditions – Two-Out-of-Three Rule and Criteria Reinforcement Exercise – on participants' JBT performance from their first to their second JBT attempt.

For sensitivity (see Figure 1), there was a significant main effect of the within-subjects factor, Time, $F(1, 219) = 159.830, p < 0.001, \eta^2 = .422$, where sensitivity significantly increased in Time 2 ($M = 1.73, SD = 0.77$) relative to Time 1 ($M = 1.05, SD = 0.49$). We also found a main effect of the between-subjects factor, Condition, $F(1, 219) = 8.166, p < .001, \eta^2 = .036$. In addition, Time and Condition had a significant interaction, $F(1, 219) = 7.068, p < .001, \eta^2 = .031$. Specifically, where the Two-Out-of-Three Rule intervention ($M_{p2} = 1.89, SD = 0.77$) showed a steeper increase in sensitivity in part 2, relative to the Criteria Reinforcement Exercise intervention ($M_{p2} = 1.553, SD = 0.73$).

For criterion bias (see Figure 2), there was a significant main effect of the within-subjects factor, Time, $F(1, 219) = 4.847, p = 0.029, \eta^2 = .022$, such that participants had a higher criterion bias in Time 1 ($M = 0.145, SD = 0.36$) compared to Time 2 ($M = 0.073, SD = 0.34$). There was no significant main effect of the between-subject factor, Condition, $F(1, 219) = 0.051, p = 0.821, \eta^2 = .0002$, and no significant interaction between Time and Condition, $F(1, 219) = 0.243, p = 0.623, \eta^2 = .0$

Figure 1.
Longitudinal effect of the interventions on overall sensitivity.

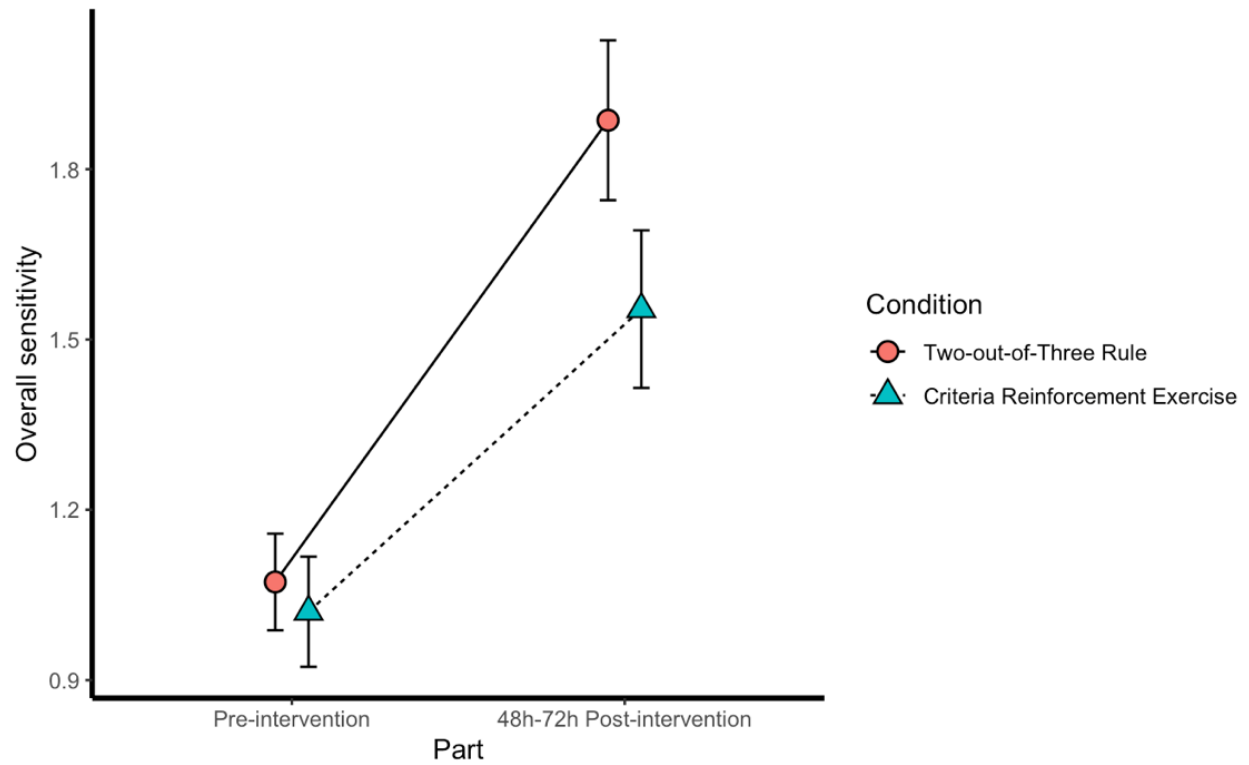
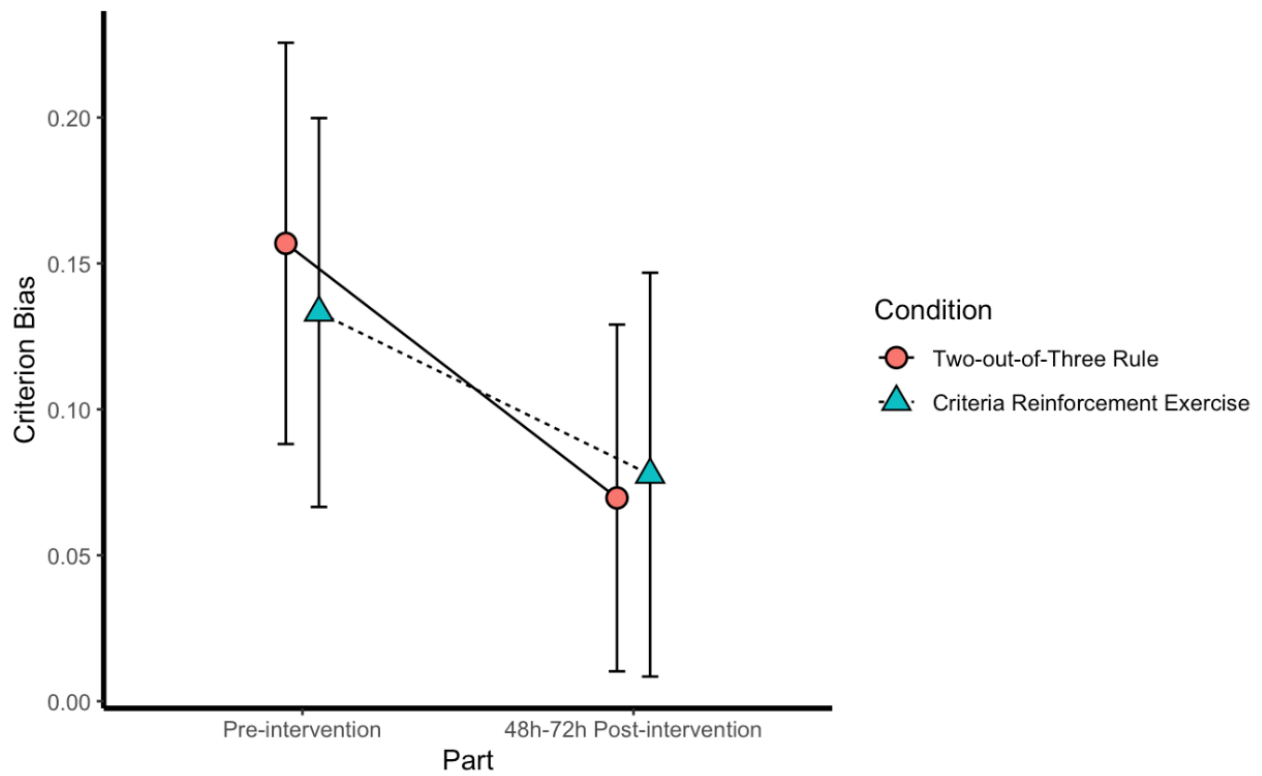


Figure 2.
Longitudinal effect of the interventions on criterion bias.



General Discussion

In this thesis, we sought to determine whether two previously successful discrimination interventions (Two-Out-of-Three Rule and Criteria Reinforcement Exercise; Roy et al., under review) would maintain their effects outside of a single study session. In a pilot study, we showed that there was no practice effect when participants completed the JBT twice in the same study session, ruling out practice effects as a confound in the longitudinal design of Study 1. This finding served to improve the validity of the results that would be found during the main part of the study. Next, Study 1 showed that both interventions successfully increased sensitivity and reduced criterion bias over a longitudinal period. In addition, the interaction between time and condition indicated that the Two-Out-of-Three Rule intervention had a significantly greater increase in sensitivity post-intervention relative to the Criteria Reinforcement Exercise intervention. However, no such interaction was found for criterion, suggesting that both interventions reduced criterion bias similarly between the first and second JBT attempts.

These findings speak to the existing research literature by emphasizing the effectiveness of behavioral plan interventions in reducing discrimination (Forscher et al., 2019; Lai et al., 2013). Implementation intentions, or if-then plans of action, have been shown to improve accuracy in tasks that measure implicit racial stereotyping (Mendoza et al., 2010; Stewart & Payne, 2008) by limiting the behavioral expression of prejudice. Both interventions tested in this study emphasize that if applicants pass a specific criterion (see procedures above for greater detail) then they must be accepted. As such, we demonstrate that the implementation interventions are also successful at reducing attractiveness-based discrimination. In addition, rubrics with more clearly defined grading criteria have been found to be less subject to discrimination for race (Quinn, 2020) and gender (Uhlmann & Cohen, 2005). These findings

provide a potential explanation as to why the Two-Out-of-Three Rule was more effective at increasing sensitivity compared to the Criteria Reinforcement Exercise. In the Two-Out-of-Three-Rule condition, participants were instructed to disregard humanities GPA and were given applicant criteria averages that had been rounded up or down. On the other hand, the Criteria Reinforcement Exercise condition gave participants the exact averages of all applicant criteria and were told to weigh all four components equally. Therefore, the former intervention had fewer selection criteria that were easier to remember than the latter, resulting in more accurate information processing. These results also relate to findings from the literature on the use of heuristics in decision-making tasks: when it comes to complex decision-making that relies on multiple dimensions of information processing, using a simple pre-established rule allows for a more accurate predictor of decisions (Dhimi, 2002). Lastly, these results align with previous the JBT studies which indicated that the identified interventions had stronger effects on sensitivity than criterion bias (Roy et al., under review).

These results provide novel insight into how behaviour can be successfully changed in the long term. In previous work, many interventions that were successful at providing short-term change in implicit racial attitudes were found to have no effect when following up longitudinally (Lai et al., 2016). In addition, a meta-analysis by Forscher et al. (2016) found that, of 22 longitudinal studies of implicit bias reduction, only 9 found significant results. A key element that may explain this divergence from the literature is that the Two-Out-of-Three Rule and Criteria Reinforcement Exercise interventions do not seek to change participants' views or attitudes towards more versus less attractive applicants, rather they seek to change the behaviour that results from the decision-making task. The relatively easy-to-remember criteria allowed participants to retain the interventions longitudinally as compared to more cognitively

demanding interventions, such as exposure to counter stereotypical exemplars, appeals to egalitarian values, and evaluative conditioning (Lai et al., 2016).

These findings speak to real-world situations as almost every work or school-related field involves a certain level of decision-making that has the potential to be influenced by socially irrelevant information. Learning how to emphasize the decision-relevant information, and maintaining these effects over time, is vital to reducing discrimination. As it is nearly impossible to eliminate decision-irrelevant information from these evaluation contexts, focusing on clearly stated and pre-established criteria for admission, diagnosis, and ruling decisions may help reduce discrimination in these real-life settings. As a result, we have identified two examples of simple interventions with flexible approaches that can be applied to a variety of different contexts.

Limitations

Despite these promising results, there are several key limitations to this study that need to be addressed. First, for Study 1, we only sampled through the McGill University and the University of Florida participant pools, meaning that our sample contains only a student population. Previous research by Sears (1986) has demonstrated that students tend to be more eager to comply with authority and instructions, have stronger cognitive skills, and are quicker to change their attitudes and behaviours when compared with older adults. This indicates that the results of the interventions may be stronger with a student sample than what is representative of a general population. Furthermore, findings were obtained for a reduction in attractiveness-based bias, however, as the JBT applicants were of only white college-aged men and women, the effects of these interventions cannot yet be generalized to other discrimination domains such as race and age. Future work should focus on obtaining a more representative sample of the general

population which may be useful in parsing the effects of a student population from the effects of the interventions themselves.

In addition, while the effects of both interventions were successful longitudinally, we only tested a two-day window which is relatively short when considering their real-life applications. A study by Chang et al. (2019) found that, despite having a robust effect on self-reported attitudes the day of, a diversity training intervention on gender-based discrimination caused no behavioral changes up to 20 weeks later. Once more, the lab-based malleability of attitudes raises doubts over the possibility of a single intervention producing long-term change of said attitudes and, in turn, their resulting behaviours. As such, it is unknown how generalizable in time the effects of these interventions are. Similarly, the JBT was conducted online which allows for a more detached decision-making experience for participants, potentially making the interventions more effective. In real-life situations, such as the aforementioned employment, medical, legal, and academic contexts, decisions are often taken through talks or interviews which are highly personalized and greatly influenced by socially irrelevant information beyond attractiveness. Therefore, future research is needed to answer these questions.

Lastly, the JBT is a highly artificial and experimental decision-making context in which numerous variables, including the socially irrelevant and decision-relevant information, are under control. As such, there are objectively correct and incorrect decisions that can be made when it comes to parsing through applicants. In contrast, real-life decision-making scenarios are often much more complex and may not have a correct answer, meaning that it may be harder for the interventions to blind socially irrelevant information while making decision-relevant criteria more salient. This shift in complexity from a lab-based setting to a real-world setting may limit generalizability as the interventions take on different effects.

Conclusion

This study demonstrated the longitudinal effectiveness of two implementation intention-based discrimination interventions (Two-Out-of-Three Rule and Criteria Reinforcement Exercise; Roy et al., under review) on a decision-making task that has previously exhibited a consistent attractiveness-based bias (JBT; Axt et al., 2018). Through the interventions, participants both increased their sensitivity and reduced their criterion bias when parsing through applicant profiles. Rather than influencing participant attitudes towards more versus less attractive applications in the decision-making task, these interventions focused on reducing discriminatory behavior by providing clear instructions for participants to follow. Despite promising results, these interventions were only tested over a two-day period, therefore future research may focus on testing the longitudinal limits of these interventions. These findings are relevant as they deepen knowledge on the mechanisms behind discrimination reduction and serve as a potential base for real-world interventions in fields that are commonly influenced by social biases, such as in hiring, education, and health care.

References

- Adamovic, M. & Leibbrandt, A. (2022). A large-scale field experiment on occupational gender segregation and hiring discrimination. *Industrial Relations: A Journal of Economy and Society*, 26(1), 34-59. <https://doi.org/10.1111/irel.12318>.
- Axt, J. R., & Lai, C. K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology*, 117(1), 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>
- Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, 94(1), 8–15. <https://doi.org/10.1016/j.jpubeco.2009.11.002>
- Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, 116(16), 7778–7783. <https://doi.org/10.1073/pnas.1816076116>
- Charlesworth, T. E. S., & Banaji, M. R. (2021). Patterns of implicit and explicit attitudes II. Long- term change and stability, regardless of group membership. *American Psychologist*, 76(6), 851–869. <https://doi.org/10.1037/amp0000810>

- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314–1329. <https://doi.org/10.1037/0022-3514.83.6.1314>
- Dhami, M. K. (2003). Psychological models of professional decision making. *Psychological Science*, 14(2), 175–180. <https://doi.org/10.1111/1467-9280.01438>
- Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, 111(2), 304–341. <https://doi.org/10.1037/0033-2909.111.2.304>
- Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, 33(1), 45–61. <https://doi.org/10.1080/02602930601122803>
- Fibbi, R., Midtbøen, A. H., & Simon, P. (2021). Concepts of discrimination. *Migration and Discrimination: IMISCOE Short Reader* (pp. 13–20). Springer International Publishing. https://doi.org/10.1007/978-3-030-67281-2_2
- 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>
- Gaddis, S.M. (2015). Discrimination in the credential society: an audit study of race and college selectivity in the labor market. *Social Forces*, 93(4), 1451–1479. <https://doi.org/10.1093/sf/sou111>.

- Gordon, R., Crosnoe, R., & Wang, X. (2013). Physical Attractiveness and the Accumulation of Social and Human Capital in Adolescence and Young Adulthood: Assets and Distractions. *Monographs of the Society for Research in Child Development*, 78, 1–137. <https://doi.org/10.1002/mono.12060>
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. *John Wiley*.
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y., J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., n Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M., Burns, M., ... Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, 145(8), 1001–1016. <https://doi.org/10.1037/xge0000179>
- Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of Attitudes or Malleability of the IAT? *Journal of Experimental Social Psychology*, 46(2), 286–298. <https://doi.org/10.1016/j.jesp.2009.11.011>
- Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, 7(5), 315–330. <https://doi.org/10.1111/spc3.12023>
- Lippens, L., Vermeiren, S., & Baert, S. (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review*, 151, 104315. <https://doi.org/10.1016/j.euroecorev.2022.104315>
- 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>

Mobius, M. M., & Rosenblat, T. S. (2006). Why Beauty Matters. *American Economic Review*, 96(1), 222–235. <https://doi.org/10.1257/000282806776157515>

Monk, E. P., Esposito, M. H., & Lee, H. (2021). Beholding Inequality: Race, Gender, and Returns to Physical Attractiveness in the United States. *American Journal of Sociology*, 127(1), 194–241. <https://doi.org/10.1086/715141>

Nault, K. A., Pitesa, M., & Thau, S. (2020). The Attractiveness Advantage At Work: A Cross-Disciplinary Integrative Review. *Academy of Management Annals*, 14(2), 1103–1139. <https://doi.org/10.5465/annals.2018.0134>

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2012). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, 15(4), 152–159. <https://doi.org/10.1016/j.tics.2011.01.005>

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, 72(1), 533–560. <https://doi.org/10.1146/annurev-psych-071620-030619>

Quinn, D. M. (2020). Experimental Evidence on Teachers' Racial Bias in Student Evaluation: The Role of Grading Scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392. <https://doi.org/10.3102/0162373720932188>

- Roy, E., Jaeger, B., Evans, A. M., Turetsky, K. M., O'Shea, B., Peterson, M. B., Singh, B., Correll, J., Zheng, D., Brown, K. W., Kirgios, E. L., Chang, L. W., Chang, E. H., Steele, J., Sebastien, J., Sedgwick, J., Hackney, A., Cook, R., Yang, X., ... Axt, J. R. (Under review). A contest study to reduce attractiveness-based discrimination in social judgment.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51(3), 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>
- Scholz, J. K., & Sicinski, K. (2015). Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study. *The Review of Economics and Statistics*, 97(1), 14–28. https://doi.org/10.1162/REST_a_00435
- Stewart, B. D., & Payne, B. K. (2008). Bringing Automatic Stereotyping Under Control: Implementation Intentions as Efficient Means of Thought Control. *Personality and Social Psychology Bulletin*, 34(10), 1332–1345. <https://doi.org/10.1177/0146167208321269>
- 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), e0148284. <https://doi.org/10.1371/journal.pone.0148284>
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed Criteria: Redefining Merit to Justify Discrimination. *Psychological Science*, 16(6), 474-480. <https://doi.org/10.1111/j.0956-7976.2005.01559.x>

Wallaert, M., Ward, A., & Mann, T. (2010). Explicit control of implicit responses: Simple directives can alter IAT performance. *Social Psychology*, 41, 152–157.

<https://doi.org/10.1027/1864-9335/a000022>

Williams, D. R., Lawrence, J. A., Davis, B. A., & Vu, C. (2019). Understanding how discrimination can affect health. *Health Services Research*, 54(S2), 1374–1388.

<https://doi.org/10.1111/1475-6773.13222>