The Impact of Timing on Reducing Bias Based on Physical Attractiveness

Tara Quinlan

Supervised by: Dr. Jordan Axt
McGill University

Course: PSYC 498D

According to the Canadian Human Rights Act (R.S.C., 1985, c. H-6), there are many grounds on which discrimination is prohibited. This includes, but is not limited to, discrimination based on race, national or ethnic origin, color, religion, age, sex, sexual orientation, gender identity or expression, or marital status. Notably however, physical attractiveness is not mentioned as a form of prohibited discrimination in Canada.

Unfortunately, appearance-based discrimination is a relatively underexplored issue, despite studies showing that even 100 milliseconds of exposure to a face is enough for people to encode physical attractiveness (Batres & Shiramizu, 2023). Moreover, human beings are biased by attractiveness, a finding that has been referred to in previous literature as the halo effect (Nisbett & Wilson, 1977), or the 'what is beautiful is good' effect (Dion et al., 1972; Eagly et al., 1991). The halo effect is the tendency that people have to attribute socially desirable personality traits to physically attractive individuals; for example, more physically attractive individuals are consistently judged to be more competent, altruistic, friendly and sincere (Batres & Shiramizu, 2023; Dion et al., 1972; Eagly et al., 1991; Nisbett & Wilson, 1977). This halo effect has been found across cultures, with one analysis (Batres & Shiramizu, 2023) finding evidence for a halo effect across 45 countries and 11 world regions.

These findings have important real-world implications. For example, studies of jury decision-making have found that less physically attractive defendants are more likely to be found guilty by mock jurors than physically attractive ones charged with the same crime (Mazzella & Feingold, 1994; Smith & Hed, 1979). Elsewhere, in hiring contexts, a Romanian study (Chelcea & Ivan., 2016) found that physical attractiveness was one of the most important factors in job recruitment, as attractive individuals were more likely to 1) end up in more senior positions within an organization, 2) get hired and promoted more quickly than average looking individuals,

and 3) negotiate higher salaries. Separate research supports these general findings that attractive people receive favorable treatment in many spheres of life including at work. For example, a more physically attractive applicant will consistently be chosen over a less physically attractive counterpart that is equally qualified (and sometimes even more qualified; Axt et al., 2018; Hammer, 2017; Maheshwari, 2022)

To better understand this issue, psychologists have developed multiple measures for studying attractiveness biases in behavior, one of which is the Judgment Bias Task (JBT; Axt et al., 2018). In the JBT, participants are instructed to assess applicant profiles of varying quality for a certain outcome based on objective and relevant decision-making criteria. For example, in an academic JBT, participants are instructed to accept about half of the applications they are presented for an honors society. The relevant decision-making criteria in this case are the applicant profiles' qualifications, such as GPA and strength of recommendation letters. What makes the task interesting is the addition of irrelevant and potentially biasing social information, (e.g., a face communicating physical attractiveness) that may undermine the use of the more objective and decision-relevant judgment criteria during evaluation.

The JBT has been able to reveal social biases in judgment that exist outside of conscious awareness or conscious control, such as biases favoring more physically attractive individuals (Axt et al., 2018), as pro-attractive biases in judgment were evident even among participants who reported afterwards that they either did not intend to use physical attractiveness in their decision-making or actually did not use physical attractiveness when making judgments.

Performance on the JBT is typically analyzed using Signal Detection Theory (SDT; Green & Swets, 1966; Axt & Lai, 2019). Specifically, SDT can be applied to participants' decisions through two outcomes : sensitivity (d') and criterion (c).

Sensitivity measures the extent to which a participant distinguishes between the two populations present in the JBT (e.g., between more and less qualified profiles). Sensitivity is then related to the amount of noise in judgment. Put simply, noise refers to the number of errors made by participants when trying to distinguish between underlying populations. Participants with high sensitivity (and less noise) are better at parsing applicant profiles and therefore commit fewer judgment errors. A score of zero in sensitivity ($d' = 0$) indicates that a participant has no ability in distinguishing more from less qualified profiles (i.e., chance levels of performance).

Alternatively, criterion measures the extent to which a participant is lenient or strict in evaluation and whether one response (e.g. to accept or to reject) is more likely. In the context of the JBT, lower criterion values indicate more leniency in evaluation; for instance, a lower criterion value would suggest that participants do not need to be as qualified in order to receive an "accept" response, and bias is present when criterion differs between targets from different social groups (e.g., lower criterion for more physically attractive applicants compared to less physically attractive applicants). In other words, while sensitivity is connected to the total amount of errors made in judgment, criterion bias details which *types* of errors are relatively more common (i.e., errors where evaluators are too lenient or too strict). As a result, discrimination is present in the JBT when there are 1) errors in judgment being made (as indexed by sensitivity) and 2) differences in criterion between targets from different social groups (as indexed by reliable differences in criterion).

The Present Work

It is important to study when social information shapes judgment, not just how it does. The current study explores this issue by examining whether attractiveness-based favoritism in judgment can be disrupted by first forcing participants' attention to more decision-relevant

information, seeing as previous research has established that initially presenting faces for as little as 350 ms at the start of a JBT trial is enough to create a criterion bias (Axt & Johnson, 2021; Roy et al., 2023).

These prior findings on timing and criterion bias came from research carried out by Roy et al. (2023) where the research manipulated the presence of decision-irrelevant information (i.e., a face communicating physical attractiveness) and decision-relevant information (i.e., candidate qualifications) during a JBT. In one study, applicants' faces would disappear after various amounts of time across conditions, while qualifications remained on screen for as long as participants needed to arrive at a decision. One condition included a timeframe in which the presentation was so short that faces were unlikely to be encoded (50 ms), while other conditions presented faces for a relatively brief period of time, but still long enough that faces could be reliably encoded (e.g., Klatt et al., 2016). Results found that, except for the 50 ms condition, the length of exposure to applicant faces did not impact sensitivity or criterion bias. In other words, a JBT that initially presented faces for as little as 350 milliseconds before removing was no different than a version that presented faces for the entirety of judgment.

The present study's goal is to further this line of research by using a reverse approach to that used by Roy et al. (2023). Specifically, instead of examining what happens when decision-irrelevant information (i.e., the face) is removed after a certain timeframe, this study investigates whether decision-making can be changed by *delaying* the presentation of the biasing social information.

Participants will be assigned to one of three conditions that manipulates the presentation format of the JBT: a control condition (where faces and qualifications are present throughout)

and two delay conditions, where faces appear either 1000 ms or 2500 ms after the relevant information (i.e., academic qualifications) first appear on screen.

These conditions were decided partly based on prior research using the JBT that showed that 500-1000 ms was insufficient of a delay between the appearance of qualifications and the appearance of a face communicating physical attractiveness to see a reduction in judgment bias (Axt & Johnson, 2021). It could be argued that were the face presented after the average time a participant takes to look at and encode the qualifications (2019 ms based on an eye-tracker study reported in Roy et al., 2023), then the average participant would have already made up their mind when the face appeared, thus potentially removing reliance on the face when making judgments.

Our hypothesis is that presentation time will impact JBT performance. Specifically, delaying the presentation of the face may force participants to give greater attention to the qualifications, which should in turn increase overall sensitivity. As for criterion biases, our hypothesis is that we will see a smaller difference in criterion bias in the 2500 ms condition because participants will have mostly finished encoding and processing the qualifications once the face does appear, compared to the 1000 ms and control conditions.

Methods

Participants. We sought to collect a minimum of 432 participants across conditions to have at least 80% power at detecting a small to medium within-subjects effect size of Cohen's *d* = 0.15. Participants were excluded if they a) failed an attention check item, b) accepted more than 80% or less than 20% of JBT applicants, or c) accepted or rejected every more or less physically attractive applicant (Axt et al., 2018). We were able to collect more data than anticipated, and we ended up with 961 participants total from the Project Implicit research pool, 850 of them being eligible following the above exclusions (M_{age} = 34.0, SD = 15.0; 60.9% White, 64.4% women).

No analyses were completed until data collection was finished. See https://osf.io/gz58u for pre-registration of study materials and analyses.

Design. Participants were randomly assigned in one of three conditions : 1) the control condition, where both the qualifications and the faces communicating physical attractiveness stayed on screen the entire judgment, 2) the 1000ms condition, where the face communicating physical attractiveness appeared 1000ms after the qualifications, or 3) the 2500ms condition, where the face appeared 2500ms after the qualifications.

Procedure. Participants first completed an academic version of the JBT, followed by a survey about JBT performance, then measures of explicit and implicit attractiveness attitudes in a randomized order, though these attitude measures were not included in any pre-registered primary analyses. Participants' demographic information was collected when initially registering for the Project Implicit research pool.

Academic decision-making task. For the JBT, participants were asked to evaluate applications for an academic honors society. They were instructed to accept the most qualified applicants and reject the least qualified and were told to accept approximately half of the applicants.

Before judging the applicants, there was an encoding phase where the participants were exposed to all 64 applications. These applications were presented one at a time for one second each, with the purpose of giving participants a sense of the range of credentials so they could make informed decisions to accept about half and reject about half of the actual applicants. For this encoding phase, no faces appeared with the profiles.

Each JBT application had four pieces of information, 1) Science GPA: GPA (out of 4.0) in science classes (biology, chemistry, etc.), 2) Humanities GPA: GPA (out of 4.0) in humanities

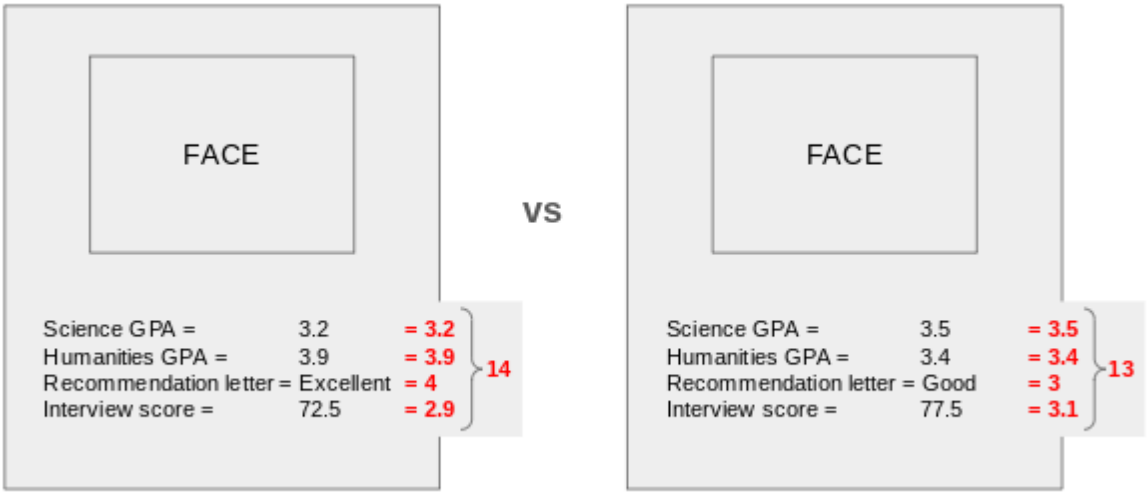
classes (English, foreign languages, etc.), 3) Letters of Recommendation: the overall quality (poor, fair, good or excellent) of the letters of recommendation in the application, and finally 4) Interview Score: the interview score (out of 100) that the applicant received during the initial round of the application process. Participants were also told that this academic honor society was looking for well-rounded applicants, so they should consider each of the four qualifications equally when they made their decision.

Sixty-four unique applicants were created, half of which were objectively more qualified and half of which were objectively less qualified. To do so, each qualification was converted to a 1-4 score— the two GPAs were already scored with a maximum of four, recommendation letters were scored either Poor = 1, Fair = 2, Good = 3, or Excellent = 4, and interview scores were divided by 25. Profiles were created such that more qualified profiles had their credentials adding up to a total score of 14 and less qualified profiles had credentials adding up to a total of 13. See Figure 1 for sample profiles for a more and less qualified applicant.

Each application profile was paired with a face that was previously pretested to be either more or less physically attractive ($d > 2.00$, Axt et al., 2018), such that both the more qualified and less qualified profile types had profiles paired with 16 more physically attractive faces (eight men, eight women) and 16 less physically attractive faces (again, eight men, eight women). Participants were assigned to one of twelve JBT orders. Across orders, each profile was equally to be paired with a more or less physically attractive face and each face was equally likely to be paired with a more or less qualified profile.

Participants were randomly assigned to one of three conditions. Conditions varied on timepoint at which they presented applicant faces during the evaluation faces (i.e., at what point in the trial faces appeared). In the control condition, the faces appeared alongside the

Figure 1. Example of a More Qualified and a Less Qualified Applicant Profile



Sample JBT applications. The application on the left is an example of an objectively more qualified profile, as the qualifications amount to a total of 14. The application on the right is an example of an objectively less qualified profile, as the qualifications amount to a total of 13. See Axt et al. (2018) for more information about the JBT.

qualifications for the entire judgment. In the 1000ms condition the face appeared 1000ms after the qualifications did. In the 2500ms condition, the face appeared 2500ms after.

1000 ms was chosen firstly because it is the longest delay tested in prior research (Axt & Johnson, 2021), which found that a 1000ms delay for presenting faces on the JBT had no impact on criterion biases relative to a control condition. Additionally, this timeframe was chosen because we wanted one condition where the face appeared before the average participant had finished looking at and encoding the relevant information (i.e., attending to applicant qualifications), as a prior eye-tracking study found that participants on average looked at qualifications for 2019ms while completing a similar JBT (Roy et al., 2023). We then chose

2500ms as the delay in our final condition because it meant the face would arrive at a time when most participants had finished looking at the qualifications.

Perceived and desired task performance. After the JBT, participants answered two items about task performance.

Participants first reported on how they perceived they performed on the task, using a seven-point scale ranging 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), with a neutral response of "I treated both physically unattractive and physically attractive applicants equally" (0).

Participants then reported on how they wanted to perform on the task. The seven-point scale used was very similar to the first, only substituting the word 'was' for 'wanted'. For example : "I wanted to be extremely easier on physically unattractive applicants and extremely tougher on physically attractive applicants" (-3).

Explicit attractiveness attitudes. Participants were asked to report their preference for physically attractive and unattractive people using a seven-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), and a neutral response of "I like physically unattractive and physically attractive people equally" (0).

Implicit attractiveness attitudes. Finally, participants were asked to complete a Brief Implicit Association Test (BIAT : Sriram & Greenwald, 2009) where they categorized more or less attractive people as well as Good words (love, pleasant, great, wonderful) and Bad words (hate, unpleasant, awful, terrible). The task assesses associations between concepts by measuring

how quickly a person can categorize, for example, positive words with relatively attractive faces compared to positive words with relatively less attractive faces. The idea is that the more strongly associated the two concepts are in memory, the more quickly participants will be able to categorize words into those paired categories.

Participants' BIAT scores are reported as an implicit preference for more physically attractive people compared to less physically attractive people if they were faster at categorizing positive words with more attractive faces compared to less attractive faces. The BIAT was scored using the *D* algorithm recommended by Nosek et al. (2014). A positive *D* score indicates faster responding on average when more attractive faces were paired with good words and less attractive faces were paired with bad words, compared with the reverse.

Results

Consistent with prior JBT studies (Axt et al., 2018; Roy et al., 2023), participants accepted nearly 50% of applicants (M = 53.1%, SD = 13.1%). Accuracy was above chance (M = 66.5%, SD = 8.8%), and sensitivity (d') was above zero (M = 1.00, SD = .55). Also replicating past work (e.g., Axt et Lai, 2019), a paired samples *t*-test showed that participants in the control condition showed a bias in response criterion, with more attractive applicants (M = -0.16, SD = .48) receiving a lower criterion than less attractive applicants (M = -0.02, SD = .48; $t(273) = 6.17$, $p < .001$, $d = 0.37$, 95% C.I. [0.25, 0.50]), meaning participants in the control condition showed greater leniency towards more versus less attractive applicants. Also, a one sample *t*-test demonstrated that participants in the control condition achieved above-chance responding ($t(273) = 29.45$, $p < .001$, $d = 1.78$, 95% C.I. [1.59, 1.97]).

Next, we conducted one-way ANOVAs across experimental conditions for sensitivity and criterion bias (a difference score between the more and less physically attractive criterion, with

higher values meaning more leniency towards more versus less physically attractive applicants). We found no main effect of condition on sensitivity ($F(2, 853) = 0.58$, $p = .944$, $\eta^2 = .0001$), nor on criterion bias ($F(2, 853) = 6.95$, $p = 0.499$, $\eta^2 = .002$). Finally, follow-up analyses found that there was a reliable criterion bias in each condition. In the 1000ms condition, participants showed a bias in response criterion in favor of the more attractive applicants ($t(288) = 4.56$, $p < .001$, $d = 0.27$, 95% C.I. [0.15, 0.39]), as did participants in the 2500ms condition ($t(292) = 4.25$, $p < .001$, $d = 0.25$, 95% C.I. [0.13, 0.37]).

See Table 1 for sample sizes and descriptive statistics within each condition for demographic information.

Table 1. Means and Standard Deviations for Criterion Bias and Sensitivity

	Criterion Bias			Overall Sensitivity		
Condition	**0**	**1000**	**2500**	**0**	**1000**	**2500**
Valid	274	289	293	274	289	293
Mean (Std. Deviation)	0.17 (0.45)	0.13 (0.48)	0.12 (0.49)	0.95 (0.54)	0.97 (0.55)	0.96 (0.56)

Discussion

The aim of our research was to investigate whether delaying the presentation of social information (i.e., a face communicating physical attractiveness) could reduce discrimination in social judgment. However, presentation delays of 1000 ms or 2500 ms did not reliably change sensitivity (i.e., errors made) or biases in criterion (i.e., greater leniency towards more versus less physically attractive people).

At the same time, our control condition findings are consistent with past work— here, participants completing a JBT in an academic context showed greater leniency towards more

attractive applicants versus less attractive ones (a criterion bias) and they achieved above-chance responding (i.e., sensitivity). This was the case for both of the experimental conditions as well, as both delay conditions produced above-chance responding in terms of accuracy and more favorable treatment of more physically attractive applicants.

These findings are also consistent with past work by Roy et al. (2023), which found that although the number of errors participants made was related to the time spent looking at the qualifications, the time spent looking at the qualifications did not affect how strict or lenient a participant was at sorting the applications depending on the applicant's attractiveness. The study by Roy et al. (2023) used eye-tracking analyses that allowed the researchers to track the amount of time participants spent attending to the different pieces of information (faces vs qualifications). Results showed that noise was consistently related to the amount of time participants viewed decision-relevant information (i.e., the more time they spent looking at the qualifications, the less errors they made) but that bias was *unrelated* to the amount of time participants spent viewing the faces (did not affect leniency). This latter result appears to be consistent with the present findings: since bias was unrelated to the amount of time spent viewing faces, it seems like any presentation of the faces (even well after the decision-making process has started) leads to attractiveness-based biases in judgment.

Current research using the JBT– and work on attractiveness-based discrimination more broadly– suggest that discrimination will persist in judgment unless participants are provided with an intervention that gives a strategy for counteracting the biasing influence of a person's face. Our research did not find any change in response bias when faces appeared 2500 ms after the qualifications were first shown, which lends further support to the idea that as soon as social information is presented, it could impact the decision-making process. Future research may want

to test additional response delays (e.g., 3500 ms or 5000 ms), but the present data is pessimistic that even these longer delays will substantively change behavior.

This work does have several limitations. For one, our sample consisted of 64% of women. This could be a further limitation of the sample seeing as it does not reflect actual population proportions which are around 50% male, 50% female according to Statistics Canada. The sample limits the ability to generalize our findings to men. This has potential implications given that men are consistently in higher-up positions and thus have more of a control in many hiring processes (O'Connor, 2019). The fact that our ability to generalize our findings to men in the context of hiring is limiting. It would be interesting to see if a more gender-representative sample shows similar patterns of findings or if the biases are more pronounced. Prior works on the subject has had similar sample demographics, in which women are overrepresented compared to population proportions (e.g., Axt & Lai, 2019; Roy et al., 2023) and even more so compared to proportions of people in powerful positions (O'Connor, 2019). That said, the proportion of participants in Studies 2a-2b by Roy et al. (2023) were only 47.9% female, which is closer to population proportions. Results found similar findings to prior research using the JBT and were similar to the results of the current study as well. This points to the assumption that men and women hold similar amounts of bias.

Not only are men in more powerful positions in companies, but there's also important elements of intersectionality to consider— the gender-based biases in hiring for example. Whereas the JBT is a quick decision making task that relies on participants spending around two seconds on judging whether to accept or reject an applicant, hiring decisions, for example, take a lot more time than two seconds to make, leaving the possibility of interplay between more than one discriminatory bias (e.g., such as being an attractive— or less attractive— woman vs man).

Prior research has found that there is gender bias in hiring decisions, usually favoring men (Hoover et al., 2019; Shields, 2008). Moreover, a study on the effects of physical attractiveness and gender on selection decisions found that while more physically attractive men were rated as most qualified and most likely to be hired, attractive women suffered a penalty in these same outcomes relative to both men and less attractive women (Shahani-Denning et al., 2011). Seeing as similar gender gaps have also been found in academia, with gaps in leadership, salary, funding and resources (Veelen & Derks, 2022), future work should look for whether biases emerge across target genders or differ across contexts (e.g., hiring for a job versus admitting to an honor society).

It would also be important to consider the intersectionality of bias linked to physical attractiveness and Eurocentric standards of beauty (i.e., race/ethnicity component). For example, African-Americans are an ethnoracial minority in Canada according to Statistics Canada (2024). African-Americans are less likely to have facial features, hair types, or skin tones that correspond with the Eurocentric standards of beauty, which could lead them to be perceived as less attractive than members of dominant ethnoracial groups (Monk et al., 2021). The fact that our sample was about 60% Caucasian and that the applicant faces were all Caucasian limits the external validity of the study by not including a diversity of race/ethnicity that would be found in real world settings such as hiring or academia. Indeed, as Monk et al. (2021) explain it, whereas Black men may face “double jeopardy” on the labor market because of the interplay of race and physical attractiveness, Black women may have to face “triple jeopardy” because of the interplay of race, physical attractiveness *and* gender. In addition, a study by Hill (2002) showed that skin tone influences physical attractiveness ratings in Black women, which suggested a pervasiveness of Eurocentric standards of beauty among African-Americans. Differences in how individuals are

perceived based on the interplay of gender and attractiveness is not only found in Black individuals, but also in other minority groups. A study by Stephen et al., (2018) found that while White men are perceived as more attractive than Asian men, White women are perceived as *less* attractive than Asian women. The above findings point to the fact that future work could use a more diverse stimulus set, which better simulate and reveal real-world interplays of gender, race/ethnicity and physical attractiveness.

A future direction to continue this line of research could be to include ethnic minorities in the applicant profiles. For example, we could have more and less physically attractive Caucasian and Black applicants by either doubling the length of the JBT (i.e., 128 applicants in all; 16 more physically attractive Black applicants, 8 of them being male, the other 8 female, etc.) or by using the same length (i.e., 64 applicants) and fewer applicants in each category.

Moreover, while the JBT is a useful tool in assessing discrimination, it is still subject to many limitations associated with lab-based measures. For example, participants in a study can act differently than they usually would, aware that they are participating in a study (i.e., reactivity bias and/or social desirability bias; Arslan et al. 2021; Fisher, 1993). In addition, participants complete the JBT on a computer, which brings forth the awareness that they are working with hypothetical scenarios. Another, perhaps more expensive way to measure discrimination based on physical attractiveness could be to construct a study utilizing confederates, immersing participants into a controlled environment in which they have to choose between candidates to admit into an honors society. This route to measuring attractiveness-based discrimination would maintain more external validity, though it would clearly require more time and effort.

Finally, even if there is internal validity with the JBT in assessing bias tied to physical attractiveness, the same can’t be said of external validity. In real-world contexts, individuals

tasked with making judgments, such as judges, juries or interviewers, must see and interact with applicants and cannot delay the potentially biasing information. And even if the potentially biasing information could be delayed, in many areas, the problem runs deeper. Simply delaying the introduction of the potentially biasing information may be insufficient, as such interventions do not provide participants with counteractive strategies to resist the influence of physical attractiveness on judgment. For example, the Google Chrome extension 'Unbias Me' is used to remove potentially biasing information from candidate profiles from GitHub, LinkedIn and Twitter such as their name and picture. This plug-in only helps in as much as the hiring practices are done on the online platforms. As soon as the candidates are met in person or over zoom, for example, the potential for bias is still there and the interviewers have no awareness or training against it, thus only delaying the problem.

There are also several other potential future directions to continue this line of research. For one, it would be interesting to further the research and investigate whether a timeframe longer than 2500 ms would show a decrease in criterion bias. For example, testing whether a delay of 4500 ms for the appearance of the decision-irrelevant information would not only significantly increase sensitivity but also decrease criterion bias. Support for this approach comes from a study by Axt & Lai (2019) using a JBT in which both decision-relevant and decision-irrelevant information were available throughout the judgment. Results found that the less time a participant is granted to pass a judgment (e.g., 1500 ms), the more errors are made (lower sensitivity). However, asking participants to deliberate about their judgment and requiring a delay of 4500 ms had the opposite effect of increasing sensitivity, though these interventions had no impact on criterion bias. A follow-up study could investigate whether a) requiring a 4500 ms delay in responding and b) presenting the biasing information after only 4500 ms could

combine to both increase sensitivity and reduce criterion bias (since even more of the judgment time would be made without the presence of attractiveness-related information).

Even when the start of the hiring process could be done with only decision-relevant information, in actual hiring practices, the individuals doing the hiring or evaluating will eventually have to come in contact with the applicants, often in an interview setting in which physical attractiveness could come into play. And so, although establishing that a longer period of time could diminish criterion bias would be interesting as it could better simulate real world processes, it would be more realistic to focus our efforts on finding other ways to reduce these forms of social judgment biases.

References

Arslan, R. C., Reitz, A. K., Driebe, J. C., Gerlach, T. M., & Penke, L. (2021). Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. *Psychological Methods, 26(2),* 175-185. doi:http://dx.doi.org/10.1037/met0000294

Axt, J. R., & Johnson, D. J. (2021). Understanding mechanisms behind discrimination using diffusion decision modeling. *Journal of Experimental Social Psychology, 95*, 104-134. doi:https://doi.org/10.1016/j.jesp.2021.104134

Axt, J. R., & Lai, C. K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology, 117(1),* 26-49. doi: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. doi:https://doi.org/10.1016/j.jesp.2018.02.011

Batres, C., & Shiramizu, V. (2023). Examining the “attractiveness halo effect” across cultures. *Current Psychology, 42(29),* 25515-25519. doi:https://doi.org/10.1007/s12144-022-03575-0

Chelcea, S., & Ivan, L. (2016). The" Aphrodite effect": Labor market discrimination based on attractiveness. *Psihologia Sociala, (37),* 37-53.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24(3),* 285-290.

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.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research, 20(2),* 303-315. doi:https://doi.org/10.1086/209351

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics (*Vol. 1, pp. 1969-2012*)*. New York: Wiley.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85(2),* 197-216. doi:10.1037/0022-3514.85.2.197

Hammer, C. (2017). A look into lookism: an evaluation of discrimination based on physical attractiveness. *Undergraduate Honors Capstone Projects*. https://digitalcommons.usu.edu/honors/207

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior, 5,* 269-298.

Hill, M. E. (2002). Skin color and the perception of attractiveness among African Americans: Does gender make a difference? *Social Psychology Quarterly, 65(1),* 77-91. doi:https://doi.org/10.2307/3090169

Hoover, A. E., Hack, T., Garcia, A. L., Goodfriend, W., & Habashi, M. M. (2019). Powerless men and agentic women: Gender bias in hiring decisions. *Sex Roles, 80,* 667-680. doi:https://doi.org/10.1007/s11199-018-0964-y

Klatt, T. Maltby, J. J., Humphries, J., Phelps, M., & Flowe, H. D. (2016). Looking bad: Inferring criminality after 100 ms. *Applied Psychology in Criminal Justice, 12 (2),* 114-125. http://www.apcj.org/journal/index.php?mode=view&item=118

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. doi: 10.1037/a0036260

Maheshwari, B. (2022). Discrimination based on physical attractiveness: Causes and consequences: A critical perspective. *Psychological Reports (Vol. 0(0))*. 1-18. doi: 10.1177/00332941221149174

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis 1. *Journal of Applied Social Psychology, 24(15),* 1315-1344. doi: https://doi.org/10.1111/j.1559-1816.1994.tb01552.x

Monk Jr, 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. doi:0002-9602/2021/12701-0005$10.00

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35(4)*, 250-256. doi: https://doi.org/10.1037/0022-3514.35.4.250

O'Connor, P. (2019). Gender imbalance in senior positions in higher education: What is the problem? What can be done? *Policy Reviews in Higher Education*, *3(1)*, 28-50. doi: 10.1080/23322969.2018.1552084

Ratliff, K. A., & Smith, C. T. (2021). Lessons from two decades of Project Implicit. *A Handbook of Research on Implicit Bias and Racism. APA Books.*

Roy et al., (2023). Disentangling the role of visual attention in discriminatory behavior. *Social Psychological and Personality Science*. [UNPUBLISHED AS OF YET]

Shahani-Denning, C., Andreoli, N., Snyder, J., Tevet, R., & Fox, S. (2011). The effects of physical attractiveness and gender on selection decisions: An experimental study. *International Journal of Management, 28(4),* 16-23.

Shields, S. A. (2008). Gender: An intersectionality perspective. *Sex Roles, 59*, 301-311. doi:10.1007/s11199-008-9501-8

Smith, E. D., & Hed, A. (1979). Effects of offenders' age and attractiveness on sentencing by mock juries. *Psychological Reports, 44(3)*, 691-694.

Stephen, I. D., Salter, D. L., Tan, K. W., Tan, C. B., & Stevenson, R. J. (2018). Sexual dimorphism and attractiveness in Asian and White faces. *Visual Cognition*, *26*(6), 442-449. https://doi.org/10.1080/13506285.2018.1475437

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56(4)*, 283-294. doi: 10.1027/1618-3169.56.4.283

Toledano, E. (2012). The looking-glass ceiling: Appearance-based discrimination in the workplace. *Cardozo JL & Gender, 19*, 683-714.

Van Veelen, R., & Derks, B. (2022). Academics as Agentic Superheroes: Female academics' lack of fit with the agentic stereotype of success limits their career advancement. *British Journal of Social Psychology, 61(3)*, 748-767. doi:10.1111/bjso.12515