

Individual differences in judgment bias: Are there reliable predictors of
biased judgment and decision-making?

Francesca Ryan 

February 27th, 2020

McGill University

Author Note

This research project was supervised by Dr. Jordan Axt.

Course number: PSYC 395

Abstract

Though discrimination based on social information is a widespread phenomenon, comparatively little is known about the psychological factors that best predict individual differences in such social judgment bias. The present study investigated whether biases in social judgment, specifically favouritism based on physical attractiveness, were related to measures of a) other forms of socially biased judgment, b) executive function, and c) objective and subjective numeracy. We found that individual differences in social judgment bias were not related to measures of executive function but were related to one other measure of socially biased judgment (faith in intuition) and were marginally related to subjective numeracy. Faith in intuition was also marginally correlated with another measure of biased judgment assessing the use of race in shoot/don't shoot decisions, suggesting that faith in intuition may be an individual difference related to a susceptibility to have social information influence judgment. These results motivate future research to elucidate whether interventions should be modified according to individual differences in faith in intuition, and further investigate the relationship between faith in intuition and judgment bias.

Keywords: social judgment bias, executive function, numeracy, faith in intuition

Individual differences in judgment bias: Are there reliable predictors of biased judgment and decision-making?

The effects of social bias – intended or unintended favouritism in evaluation, judgment or behaviour – are widespread and impactful. Important decisions related to admissions, hiring and promotion are frequently impacted by such biases (Bickel et al., 1975; Cash & Kilcullen, 1985; Olson & Becker, 1983). As a result, biases in social judgment and their impact on intergroup disparities have become a popular subject of psychological research. However, past research has found little evidence for robust predictors of individual differences in the likelihood of displaying such judgment biases. A better understanding of *who* is most likely to exhibit judgment biases may in turn aid ongoing efforts to develop effective interventions that can reduce unintended discrimination on a large scale.

Background

One of the most illustrative findings of early research on discrimination was that predicted and actual behaviour are not always aligned. Richard Lapierre's influential study on discriminatory attitudes toward Chinese people was instrumental in this discovery (LaPiere, 1934). In a survey administered to 251 businesses in the United States, 92% (of 128 respondents) said they would turn away Chinese customers. However, when Lapierre visited these establishments with a Chinese couple, they were only turned away once. This study established the possibility that large and consequential discrepancies may exist between individuals' intended and actual behaviour.

Due to these discrepancies, a notable challenge in subsequent bias research lay in inaccuracy of self-report measures, where participants often report positive intergroup attitudes and high levels of motivation to control prejudiced behaviour (e.g., Crosby, Bromley, & Saxe,

1980; Devine & Elliot, 1995). Despite these generally positive self-reported intergroup attitudes and motivations, research demonstrating that discrimination significantly affects real world outcomes is abundant. For instance, in one Swedish field study (Rooth, 2009) researchers applied to multiple real job openings with two fictitious job applications that were equated on outcome-relevant criteria like strength of past work experience but differed on physical appearance. Specifically, within each pair, the photo accompanying each applicant's face was manipulated to appear obese (in Europe, it is common to include headshots in application materials). Despite being similarly qualified, analyses found that faces appearing obese received significantly lower call-backs (6% lower for men and 8% lower for women), demonstrating the consequential effects of discrimination toward obese individuals.

Field studies are valuable in demonstrating the effects of bias in the real world in a way self-report measures cannot. However, one weakness of field studies is their inability to identify *which* individuals are most likely to exhibit socially biased judgment. Doing so would require including additional attitudinal, motivational or demographic measures that may not be feasible in most field-based studies, as doing so could be overly costly for researchers or could indicate to participants that their behaviour is being monitored.

Identifying the psychological variables that best predict biases in social judgment must then rely mostly on laboratory data. Yet one difficulty in such work is in recreating a context where discrimination may occur, as laboratory participants knowing they are in a study may be vigilant to not reveal biased behaviour (eg. Axt & Trawalter, 2017). Progress on this issue has come from a recently developed measure of judgment bias that simulates real-world situations: the Judgment Bias Task (JBT; Axt, Nguyen, & Nosek, 2018). Whereas the Rooth (2009) study evaluated real world outcomes, the JBT seeks to simulate these conditions in a lab setting. In a

JBT, participants are told to evaluate applicants for an outcome, such as admission to a hypothetical honour society. Researchers can manipulate the objective qualifications associated with these applicants as well as the social information assigned to each (for example, participants may evaluate applicants for an academic honour society based on relevant information like GPA, but researchers can also introduce an irrelevant social factor like physical attractiveness through pairing applications with faces). This design allows researchers to examine how such social information impacts individuals' ability to evaluate the objective qualifications. In the context of physical attractiveness, past studies have repeatedly found that participants are more lenient when evaluating more versus less physically attractive people, even when objective qualifications are equal between the two groups and when participants self-report a desire to behave objectively (Axt & Lai, 2019; Axt et al., 2018).

Consistently recreating judgment biases based on physical attractiveness in a laboratory context can accelerate practical and theoretical progress concerning how discrimination occurs and how it can be reduced. In particular, the present study explores what other psychological variables best predict bias on the JBT. Finding such variables could inform future efforts to develop interventions for reducing discrimination. For instance, a 12-week intervention developed by Devine et al. (2012) looking to reduce implicit racial attitudes (measured by an Implicit Association Test) found that the effectiveness of the intervention was moderated by concerns about discrimination. Participants demonstrating higher concern about discrimination at week 2 showed lower levels of implicit racial bias at weeks 4 and 8. The intervention was most effective on high-concern participants, but low-concern participants still showed reduction in implicit bias compared to control. While this study concerned altering IAT performance, a similar intervention approach could be applied to socially biased judgment on the JBT. By

finding what factors best predict discrimination, individuals could be identified who are most in need of targeted interventions for reducing discrimination.

Prior Research on Predictors of Biased Judgment

The present work investigates what factors best predict socially biased judgment. Prior research on this topic has focused on attitudes, both implicit and explicit, as well as measures of motivation to control prejudiced behaviour. These studies have found reliable but small correlations: that is, past work has failed to find any individual difference that could explain a substantial proportion of discriminatory behaviour. For example, some research has focused on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) to investigate whether the measure of implicit attitudes can predict biased behaviour. The study with the greatest external validity using this approach (Rooth, 2010) found that recruiters with more negative implicit attitudes regarding Arab-Muslim men were less likely to invite Arab-Muslim applicants for an interview ($r = .16$). Though small, this effect size is even somewhat larger than that obtained in a recent meta-analysis of over 200 research reports of IAT correlation with intergroup behaviour ($r = .10$, Kurdi et al., 2019). While the IAT does appear to reliably predict intergroup bias in behaviour, such effects remain small.

In terms of implicit and explicit attitudes predicting JBT performance specifically, prior research has similarly found reliable but small effects. For instance, Axt et al. (2018) found a weak correlation between implicit attitudes (measured with an attractiveness IAT) and JBT performance ($r = .15$), while a meta-analysis across eight studies found a similarly weak correlation ($r = .09$; Axt & Lai, 2019). Explicit attitudes regarding preferences for more versus less physically attractive people fared little better ($r = .11$; Axt & Lai, 2019). Finally, greater internal motivation to control prejudice was positively but weakly related to showing lower bias

on the JBT ($r = -.09$ and $r = -.07$ in Axt and Lai, 2019). These findings suggest that measures of implicit attitudes, explicit attitudes and motivation to control prejudice have reliable but weak predictive validity for the JBT. Although such effects may meaningfully contribute to intergroup disparities (Greenwald, Banaji, & Nosek, 2015), none of these variables could be characterized as strong predictors of the likelihood of exhibiting discrimination on the JBT.

The Present Work

The present study looks to extend this prior work by investigating a larger range of possible predictors of biased behaviour on the JBT. Specifically, we explore whether performance on the JBT is predicted by 1) other measures of socially biased judgment, 2) measures of executive function, and 3) measures of objective or subjective numeracy.

Other measures of socially biased judgment. If there exists a general individual difference in the susceptibility for behaviour to be influenced by irrelevant social information, then performance on the JBT may be related to other measures of judgment biases. To explore this possibility, the current study investigated whether JBT performance was associated with measures of individual differences concerning the use of race in identifying guns versus harmless objects (Correll, Park, Judd, & Wittenbrink, 2002), the likelihood of committing the fundamental attribution error (Scopelliti, Min, McCormick, Kassam, & Morewedge, 2016) and faith in intuitive judgments (Epstein, Pacini, Denes-Raj, & Heier, 1996).

Executive function. Some evidence suggests that more cognitive factors may be related to social biases in attitudes or behaviour. For instance, Ito et al. (2015) investigated the relationship between implicit racial bias, assessed through a First Person Shooter Task (Correll et al., 2002), a Weapons Identification Task (Payne, 2001), and executive function (EF): a set of mental skills heavily implicated in self-control (Baumeister, Schmeichel, & Vohs, 2007). Here,

the larger construct of EF was decomposed into skills concerning task switching, response inhibition and working memory updating. Results found a significant but moderate negative correlation between ‘common’ executive function, calculated by Ito et al. (2015) from nine measures of EF, and implicit bias, a latent variable constructed from performance on the WIT and FPST tasks ($r = -.23$). Such results suggest there may be underlying factors (i.e., executive function) that influence the display of social biases across contexts. Similar results may occur when using the JBT as a measure of biased judgment, as it is possible that individuals with weaker executive function may be more susceptible to having behaviour shaped by ostensibly irrelevant social information, such as physical attractiveness. In our study, we administered the Stroop task and the Letter Memory tasks, which had reliable associations (Stroop $r = .47$ Letter Memory $r = .54$) with the common EF variable used in Ito et al. (2015)

Numeracy. Finally, performance on the JBT may be related to individuals’ comfort with numeric information. The JBT requires participants to assess numeric information (e.g., through evaluating numeric scores on criteria like GPA’s and interview scores). As such, it is possible that participants less comfortable with numbers might show greater reliance on social information like attractiveness due to an inability or lack of interest in processing numbers. Notably, we included measures of both objective numeracy (i.e., actual ability in reasoning about numbers) and subjective numeracy (i.e., individual feelings about dealing with numeric information). Assessing both objective and subjective may then shed light on whether JBT performance is more associated with one’s ability to process numeric information (objective numeracy) versus one’s motivation or comfort with doing so (subjective numeracy). Taken together, these measures should provide further information on how ease with numbers is related to biases in social judgment.

Through this study, we seek to deepen our understanding of the psychological processes most related to social judgment biases. Data will shed light on whether susceptibility to bias in social judgment is an individual difference that may be related to other known psychological constructs and can reveal other social or cognitive factors that may be associated with the process of discrimination. Moreover, if bias on the JBT can be reliably predicted by other measures, the potential applications are substantial. For instance, interventions could be tuned or targeted towards those individuals most likely to exhibit bias. Similar approaches have proven successful in online marketing, where personality traits can be inferred from online footprints (e.g., Facebook likes) and marketers use that information to deploy ads that are known to be more or less effective for certain personalities (Matz, Kosinski, Nave, & Stillwell, 2017). Future research could then try to infer other characteristics, like working memory or subjective numeracy, from online behaviour and use that information to target individuals who may be most likely to show biases in social judgment, sending such individuals tailored interventions that are most effective for reducing the likelihood of discrimination¹.

Method

Participants

Eighty-three participants completed the study in exchange for either partial course credit or \$10. Six participants were excluded from analysis due to inability to complete the study within the allotted 1-hour time frame, resulting in 77 participants (84.4% female, $M_{Age} = 21.0$, $SD_{Age} = 2.49$) eligible for analysis.

¹This type of data collection can be seen as either beneficial or covertly exploitative. Used in the best possible manner, these targeted interventions could have the effect of reducing the bias on a wide scale, but the risks to privacy should first be carefully considered.

Procedure

In a random order, participants completed a JBT, two measures of executive function, three measures of other social judgment biases, and three measures of numeracy.

Judgment Bias Task. In the JBT developed by Axt et al. (2018), participants are told that they will be playing the role of a decision-maker, and that it is the participant's job to decide whether to give an accept or reject response to each applicant in an academic honour society. Applicants were shown with outcome-relevant criteria like GPA, interview score, and recommendation letter strength. The strength of applicants was calibrated to create two groups: half that, based on their aggregate ratings, should be rejected and half that should be accepted. Furthermore, half of the applicants in each 'accept' or 'reject' group were paired with faces that were previously rated to be more versus less physically attractive.

Our primary outcome of interest in this study is the difference in criterion for more vs. less physically attractive applicants. Response criterion assesses the degree to which participants are more likely to make more beneficial errors (i.e., admitting less qualified applicants) compared to more detrimental errors (i.e., rejecting more qualified applicants). Prior research has shown that in general, people have a lower criterion for (i.e. are more lenient toward) applications paired with attractive faces (Axt et al., 2018). Our primary JBT outcome was then a difference score subtracting criterion for less physically attractive applicants from criterion towards more physically attractive applicants, such that a higher criterion bias indicates greater leniency towards more versus less physically attractive applicants.

Other measures of socially biased judgment. Participants completed a First-Person Shooter Task (Correll et al., 2002), the Neglect of External Demands Scale (Scopelliti et al., 2016), and the Faith in Intuition Scale (FIS; Epstein et al., 1996).

First-Person Shooter Task (FPST). Participants completed an 80-trial FPST, the same used in Ito et al. (2015). The task started with a practice block of 16 trials. On each trial, participants see one to four background images (500-800ms duration). A target individual, either White or Black, then ‘appears’ on the final background for 590 ms, holding one of several handguns or one of several innocuous objects. Participants were instructed at the onset of the task to either press a button labelled “Shoot” for armed targets, or “Don’t Shoot” for unarmed targets as quickly as possible. Scoring procedures were based on Ito et al. (2015), as follows:

If a response was made within the 590 ms window, participants received feedback and points were added to their score:

- (a) Correct decision to shoot—“Good shot,” 10 points;
- (b) Incorrect decision to shoot—“You shot a good guy!!” 40 point penalty;
- (c) Correct decision not to shoot—“Wise choice,” 5 points; and
- (d) Incorrect decision not to shoot—“YOU’RE DEAD!!” 20 point penalty.

If participants failed to respond while the target appears on screen, they received a 50-point deduction. Parallel signal detection analysis to compute racial differences in criterion were used as in Ito et al. (2015). Our primary FPST outcome was then a difference score for criterion to indicate ‘gun’ for Black versus White targets, such that higher scores indicate a lower threshold to provide a ‘gun’ response for Black relative to White targets.

Neglect of External Demands Scale (NEDS). The NEDS is a 10-item questionnaire meant to assess individual differences in the fundamental attribution error. In developing the

scale, Scopelliti et al. (2016) evaluated its predictive validity, showing that responses on the scale reliably predicted a variety of potentially related behaviours, including the sensitivity of performance evaluations to job difficulty and consideration of a coerced confession in a mock trial as salient evidence.

A sample item from the scale is, “Mary is working on an essay on the negative aspects of capitalism that she was assigned to prepare for her sociology course. The main points she plans to include in her essay are that capitalism is inherently exploitative, that it leads to imperialism and oppression, and that it creates wasteful practices such as planned obsolescence of products. Based on the information provided, how confident are you that Mary’s attitude toward capitalism is negative?” Participants respond on a scale of “1= Not at All Confident” to “7= Very Confident”. Items were randomized and is scored from taking the average of all items, reverse scored when appropriate. See Appendix A for all items.

Faith in Intuition Scale (FIS). The FIS consists of twelve items assessing comfort with using intuition in evaluating others. Past research has suggested that higher reliance on intuitive decision making results, as measured by the FIS, is associated with biases in social evaluation, such as being associated with increased gender stereotyping following a positive mood induction (Trent & King, 2013). A sample item is: When it comes to trusting people, I can usually rely on my "gut feelings." Participants respond on a 5-point scale ranging from “1= Completely false” to “5= Completely true”. Items were randomized and averaged, as well as reverse scored when appropriate. See Appendix B for all items.

Measures of executive function. Participants completed a Stroop test and a letter memory task.

Stroop test. The method and scoring of this task was adapted from J.R. Stroop (1935). Participants were asked to place their fingers on four keys that correspond to four colours: red, green, blue and black. Participants were asked to press the key corresponding to the colour of the font of the word that appears as quickly and accurately as possible. There were two sections consisting of 25 trials: 5 repetitions each of the 5 colour words used (red, green, blue, black, yellow) presented in each of the 5 colours once. Performance was calculated using an adapted *D* score (Greenwald, Nosek, & Banaji, 2003), which took the reaction time difference between incongruent and congruent trials divided by the pooled standard deviation. Higher scores then indicated slower responses to incongruent versus congruent trials.

Letter memory task. The method and scoring of this task were adapted from Friedman et al. (2008). Participants were shown a series of 5, 7 or 9 random consonants appearing for 2500 ms each. Once the entire string had appeared, participants were asked to report the final four letters of the string. Participants were scored as correct even if the reported letters were out of order. The test involved three practice and 12 critical trials. A participant's score was the proportion of letters correctly recalled across all critical trials.

Measures of numeracy. Participants completed a Cognitive Reflection Test (Frederick, 2005) to assess objective numeracy. Participants also completed the Need for Cognition (Cacioppo, Petty, & Feng Kao, 1984) and Subjective Numeracy (Fagerlin et al., 2007) scales to assess subjective numeracy.

Cognitive Reflection Test (CRT). The CRT consists of three items testing participants' ability to override a 'prepotent' response to a seemingly simple numeric problem. Past research suggests the CRT predicts performance on heuristics-and-biases tasks, where higher CRT performance is correlated with lower rates of 'miserly' processing: the tendency to solve

problems in simpler, less effortful ways than ideal (Toplak, West, & Stanovich, 2011). Prior research also suggests performance on this task is closely related to numerical ability (Welsh, Burns, & Delfabbro, 2013). A sample item is “A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?”. Participants were scored 0, 1, 2, or 3 depending on the number of questions answered correctly.

Need for Cognition Scale (NFCS). The NFCS is a scale meant to assess comfort with and desire to think deeply. Participants are asked to rate the extent to which they agree with each statement using a -4 (Very strongly disagree) to +4 (Very strongly agree) scale. A sample item is, “I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.” There are 18 total items and scoring was computed by taking the overall average, reverse scoring when appropriate. See Appendix C for all items.

Subjective Numeracy Scale (SNS). The SNS measures individuals’ own feelings of numeric ability. This questionnaire includes 8 items. A sample item is: When people tell you the chance of something happening, do you prefer that they use words (“it rarely happens”) or numbers (“there’s a 1% chance”)?” and participants respond using a 1=Always prefer words to 6=Always prefer numbers scale. Participants’ score was the mean of their responses to all eight questions, reverse scoring when appropriate. See Appendix D for all items.

Results

See Table 1 for descriptive statistics for all study outcome measures.

Attractiveness and JBT performance

As in prior research (Axt & Lai, 2019; Axt et al., 2018), participants had a lower criterion for providing an ‘accept’ response towards more physically attractive applicants ($M = -.05$, SD

= .36) than less physically attractive applicants ($M = .14$, $SD = .38$), $t(76) = 4.22$, $p < .001$, $d = .48$.

Race and FPST performance

As in prior research (Correll et al., 2002), participants had a lower criterion to indicate a ‘gun’ response for Black targets ($M = -.14$, $SD = .30$) than White targets ($M = .01$, $SD = .28$), $t(76) = 3.88$, $p < .001$, $d = .44$.

Congruency and Stroop performance

A one-sample t -test revealed that Stroop performance D scores were reliably greater than zero, indicating faster reaction times to congruent than incongruent words, $t(76) = 11.77$, $p < .001$, $d = 1.34$.

Predicting JBT performance

We next ran a series of correlations predicting JBT criterion bias with all other study measures. See Table 2 for all correlations. The only significant correlation was between JBT criterion bias and the FIS, $r = .246$, $p = .031$. The correlation between JBT criterion bias and the Subjective Numeracy scale was marginally significant, $r = .213$, $p = .062$.

Discussion

This study investigated whether biases favouring more versus less physically attractive people were reliably correlated with measures of a) other social judgment biases; b) executive function; and c) objective and subjective numeracy. Only one measure (Faith in Intuition Scale) reliably predicted biases on the attractiveness JBT used to measure favouritism towards more versus less physically attractive people, though another measure (subjective numeracy) had a marginal correlation with JBT criterion bias.

One notable result was the lack of any association between the JBT and another measure of social biases in judgment, the FPST. The two tasks share several similarities. For instance, both tasks are scored using criterion bias and both target bias based on appearance (physical attractiveness versus race), but performance on the two tasks was unrelated ($r = .010, p = .933$). This null result could not be accounted for by a lack of biased behaviour in our sample, as analyses replicated results of lower criterion for an ‘accept’ response towards more physically attractive people in the JBT ($d = .48$) and a lower criterion for a ‘gun’ response towards Black targets in the FPST ($d = .44$).

The lack of a correlation between the JBT and FPST could suggest that the biased behaviour assessed in these measures is more situationally specific than a stable individual trait. In this case, an individual high in attractiveness bias would not be more likely to be high in racial bias, as performance on the two measures is not driven by an overarching trait concerning how susceptible an individual may be to having their judgments shaped by irrelevant social information.

However, it is also possible that the two tasks are simply too different to compare despite their surface-level similarities. For example, the FPST involves speeded decisions, while the JBT does not. As such, it’s possible that the FPST and JBT are not correlated because FPST performance is highly moderated by executive function while the JBT is not. Ito et al. (2015) determined that ‘common’ EF is significantly correlated to implicit bias as a latent variable constructed from scores on the Weapons Identification Task (WIT) and the FPST. However, we did not find significant correlations between the FPST and either the Stroop ($r = .174, p = .130$) or Letter Memory ($r = -.087, p = .453$) tasks. There was also no reliable correlation between these EF tasks and the JBT (Stroop $r = -.026, p = .824$; Letter Memory $r = .069, p = .550$). These

analyses suggest that executive function does not significantly impact results of either bias task, and as such they may be comparable in this respect.

However, the FPST and JBT correlations with subjective numeracy present contrary evidence. SNS scores were correlated with the JBT ($r = .213, p = 0.062$), but were not significantly correlated with the FPST ($r = .048, p = .677$). Though these two correlations themselves were not significantly different from each other (Williams $t = 1.03, p = .307$), the pattern of results from this initial sample suggests there may be an important difference between the FPST and JBT. While it is possible that these two measures of socially biased judgment are too different to be comparable, it is unclear whether their lack of correlation is due to their differences or due to the extent to which bias is situationally bound. Progress in investigating this question could be explored in future research that uses JBT and the FPST measures that are more similar. For example, the attractiveness JBT could be compared to a racial JBT (i.e. Axt, Ebersole, & Nosek, 2016) which would measure criterion bias favouring white applicants, controlling for attractiveness. A different approach could compare the traditional FPST to an attractiveness FPST, where participants must shoot or not shoot faces varying on attractiveness. Finally, it may be productive to adapt the measures used here to have more procedural similarities, such as in imposing a response window on the JBT (e.g., Axt & Lai, 2019). These variations in study design might produce larger correlations than what was observed in the current study, indicating that racial bias and attractiveness bias are correlated, but preliminary evidence on this question finds a perhaps surprising lack of a correlation.

Our findings also support an association between faith in intuition and criterion bias on the JBT. Prior research has linked faith in intuition with behavioural biases such as the representativeness heuristic (overweighting sample information; Alós-Ferrer & Hügelschäfer,

2012), which is a possible mechanism of biased judgment. For example, in a study by Triplet (1992), subjects read patient case descriptions suffering from one of two diseases, where patients differed only on symptoms and sexual orientation. Subjects were more likely to perceive homosexual patients as having AIDS regardless of their symptoms, demonstrating the representativeness heuristic at work in producing biased judgment. Similar heuristics may be related to attractiveness JBT performance. In a series of experiments, Monin and Oppenheimer (2005) demonstrated a ‘warm glow’ heuristic, whereby more attractive faces are more frequently mistaken as familiar. A similar effect may occur in JBT performance: for those high in FI, more attractive stimuli are viewed as more familiar, and this positivity translates into a greater likelihood of acceptance. Participants high in faith in intuition may be more dependent on familiarity in judgment, resulting in less reflection on decisions and greater criterion bias in the JBT.

Our results also provide mild support for faith in intuition as a predictor of social bias across contexts. The FIS was marginally correlated with the FPST ($r = .205, p = .073$) and positively but not reliably with the NEDS ($r = .124, p = .283$). Although future research is needed to render a more decisive verdict on these small correlations, a case can be made for faith in intuition as a predictor of biased behaviour across contexts. For the FPST, the representativeness heuristic can easily be conceptualized as a factor in formulation of racial bias. Given widespread portrayals in the media as Black people being more likely to be portrayed as criminals (Chiricos & Eschholz, 2002; Welch, 2007), participants high in faith in intuition may be more susceptible to viewing Black people as more dangerous, leading to lower criterion for ‘shoot’ decisions in the FPST. However, although the FIS may be related to the FPST, our results do not show a significant correlation with our third measure of social bias, the NEDS (r

= .124, $p = .283$). The NEDS directly measures propensity to infer stable characteristics from situational factors (i.e. sample information), so this finding is surprising, emphasizing the need for further investigation into the relation between faith in intuition and biased behaviour.

In either case, the FIS appears to be related to JBT performance, though data from the full sample will be more informative. But given these interim data, interventions seeking to reduce bias in the context of the JBT might increase effectiveness by targeting people who are high in faith in intuition. In particular, such an intervention might ask participants to first infer traits based on faces varying on attractiveness, then correct their mistakes with an emphasis on discouraging intuitive judgments. This feedback may then reduce the use of physical attractiveness in evaluating others and lessen the degree to which more physically attractive people receive favourable treatment.

Another result of interest is the *positive* correlation between the SNS and JBT criterion bias ($r = .213$, $p = 0.062$). If a participant wishes to perform flawlessly on the admissions JBT used here, they must standardize and compute the average of the four relevant criterion (Science GPA, Humanities GPA, recommendation letter strength, and interview) score, and then track whether each applicant is above or below the sample average. As such, one might expect higher competency with numbers to lead to *less* biased judgments. However, the SNS correlation obtained suggests that individuals who see themselves as more capable with numbers are in fact more susceptible to criterion bias. This may be a similar phenomenon to the faith in intuition result, where individuals' faith in their ability with numbers decreases the amount of reflection they feel is needed to perform well on the task. If this is the case, bias reduction interventions may benefit from targeting subjective numeracy with the JBT. Such an intervention could administer a faceless JBT, providing feedback on whether the participants' choice should have

been admitted (i.e. above or below average). This feedback may reduce participants' confidence in their judgments about numbers, resulting in more careful reflection on their decisions and a reduction in biased outcomes.

Limitations

The current study has several limitations. For one, recruitment of participants took place primarily within the campus and community of McGill University. As such, participants consisted largely of young adults ($M_{\text{Age}} = 21.0$) enrolled in post-secondary education. A sample limited to this subset may not have generalizable results. For instance, our sample was predominantly female (88.4%). This may have affected CRT results, as scores on the CRT were generally low (i.e., an average of just one correct answer) and males tend to have higher CRT scores (Frederick, 2005). A more diverse sample in terms of sex may have introduced more variance on the CRT, which may have better allowed us to detect a possible correlation between the CRT and JBT.

Age is another factor that might impact the relationship between biased judgments and cognition. In particular, a sample with a greater age range may have provided more variance on our measures of executive function. For instance, letter memory accuracy (89.96%) was surprisingly higher than what was observed in initial pilot studies, and these high levels of accuracy could have introduced a ceiling effect that suppressed possible correlations with other measures used in this study. Including a greater range of participant ages in future studies may then produce different results from what was observed here. In general, our study would have benefitted from a more diverse sample, not only for increasing generalizability but also for introducing variance that may have been undetected with such a homogenous sample.

Our results are additionally limited by the small sample size achieved to date. Future research will continue data collection, with the goal of obtaining 250 participants. This sample size would provide 80% power for detecting a correlation as small as $r = .175$ between any two measures. This statistical power may reveal more significant correlations in our measures, particularly the NEDS, which demonstrates an interesting but unreliable correlation of $r = .179$ with our current data. A larger sample size may provide the statistical power needed to determine whether this effect is significant. At the same time, the current sample size can at least rule out the presence of large effects (i.e., the sample has 95% power to detect a correlation of $r = .39$). If any of our measures other than FIS are correlated with JBT performance, they are likely to be relatively small.

Future Directions

As previously discussed, a salient finding of this study is the significant correlation found between faith in intuition and judgment bias as well as FPST performance. Despite the significant implications of these findings, much remains to be elucidated about the relationship between faith in intuition and biased judgment. While faith in intuition has previously been linked to behavioural biases (Alós-Ferrer & Hügelschäfer, 2012), future research could further explore the relationship between faith in intuition and discriminatory behaviour. Limited past research suggests faith in intuition may moderate the link between implicit and explicit attitudes (Zimmerman, Redker, & Gibson, 2011), but further evidence for its relation to discrimination is needed. For instance, faith in intuition may moderate the effectiveness of bias reduction interventions. It is possible that individuals who are high on faith in intuition may be less receptive to interventions, or that interventions targeting reduced reliance on intuitive judgments might be most effective for them.

Though the present study found little evidence of an individual difference measure that was strongly related to JBT performance, further investigation of possible predictors may be of value. For instance, attributional style, in particular internal attributional style, is significantly correlated with the neglect of external demands scale (Scopelliti et al., 2016). This tendency to make dispositional attributions may be related to JBT performance in that extraneous information may lead to stronger dispositional attributions. The Attributional Style Questionnaire (ASQ; Peterson et al., 1982), originally developed as an assessment of individual differences in the types of attribution characteristic of depression (internal, stable, global), would be a useful measure for this purpose, and future investigations may find a reliable association with JBT performance.

Conclusion

The current study found preliminary evidence of an association between faith in intuition and subjective numeracy with JBT performance. We found no evidence of a reliable relationship between the JBT and the other measures of social bias, which may suggest that bias is more situationally specific than a stable individual trait. However, in addition to the correlation between faith in intuition and the JBT, we found a marginal correlation between faith in intuition and the FPST, which may contribute to an underlying susceptibility to bias. These results lend some insight into which individuals are most likely to exhibit judgment biases, but more research remains to be done. While the current evidence finds little support of a robust predictor of individual differences in bias, there are several promising avenues for further investigations to explore. Continuing research in this area is vital, as the pervasive and impactful nature of bias in real world context becomes increasingly clear. A better understanding of which individuals are

most likely to exhibit judgment biases could prove vital in developing more successful interventions and working toward a world with less biased outcomes.

References

- Alós-Ferrer, C., & Hügelschäfer, S. (2012). Faith in intuition and behavioural biases. *Journal of Economic Behaviour & Organization*, *84*(1), 182-192. doi:10.1016/j.jebo.2012.08.004
- Axt, J., Ebersole, C., & Nosek, B. (2016). An unintentional, robust, and replicable pro-Black bias in social judgment. *Social Cognition*, *34*, 1-39. doi:10.1521/soco.2016.34.1.1
- 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:10.1016/j.jesp.2018.02.011
- Axt, J. R., & Trawalter, S. (2017). Whites demonstrate anti-Black associations but do not reinforce them. *Journal of Experimental Social Psychology*, *70*, 8-18.
- Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. In *Social psychology: Handbook of basic principles, 2nd ed.* (pp. 516-539). New York, NY, US: The Guilford Press.
- Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, *187*(4175), 398-404.
- Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.
doi:10.1207/s15327752jpa4803_13
- Cash, T. F., & Kilcullen, R. N. (1985). The eye of the beholder: Susceptibility to sexism and beautyism in the evaluation of managerial applicants. *Journal of Applied Social Psychology*, *15*(7), 591-605. doi:10.1111/j.1559-1816.1985.tb00903.x

- Chiricos, T., & Eschholz, S. (2002). The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. *Journal of Research in Crime and Delinquency*, 39(4), 400-420. doi:10.1177/002242702237286
- 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. doi:10.1037/0022-3514.83.6.1314
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, 87(3), 546-563. doi:10.1037/0033-2909.87.3.546
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton Trilogy revisited. *Personality and Social Psychology Bulletin*, 21(11), 1139-1150. doi:10.1177/01461672952111002
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 1267-1278.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390.
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making*, 27(5), 672-680. doi:10.1177/0272989x07304449
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008).

Individual differences in executive functions are almost entirely genetic in origin. *J Exp Psychol Gen*, *137*(2), 201-225. doi:10.1037/0096-3445.137.2.201

Greenwald, McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in

implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464-1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the

Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561.

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

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., &

Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187-218. doi:10.1037/a0038557

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., . . . Banaji, M.

R. (2019). Relationship between the Implicit Association Test and intergroup behaviour: A meta-analysis. *American Psychologist*, *74*(5), 569-586. doi:10.1037/amp0000364

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces*, *13*(2), 230-237. doi:10.2307/2570339

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an

effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714-12719. doi:10.1073/pnas.1710966114

- Monin, B., & Oppenheimer, D. M. (2005). Correlated Averages vs. Averaged Correlations: Demonstrating the Warm Glow Heuristic Beyond Aggregation. *Social Cognition, 23*(3), 257-278. doi:10.1521/soco.2005.23.3.257
- Olson, C. A., & Becker, B. E. (1983). Sex discrimination in the promotion process. *ILR Review, 36*(4), 624-641. doi:10.1177/001979398303600407
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*(2), 181-192. doi:10.1037/0022-3514.81.2.181
- Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The attributional style questionnaire. *Cognitive Therapy and Research, 6*(3), 287-299.
- Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of human resources, 44*(3), 710-735.
- Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics, 17*(3), 523-534. doi:10.1016/j.labeco.2009.04.005
- Scopelliti, I., Min, H. L., McCormick, E., Kassam, K., & Morewedge, C. (2016). Individual differences in correspondence bias: Measurement, consequences and correction of biased interpersonal attributions. *Management Science, 64*. doi:10.1287/mnsc.2016.2668
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi:10.1037/h0054651
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition, 39*(7), 1275.

- Trent, J., & King, L. A. (2013). Faith in Intuition moderates the effects of positive affect on gender stereotyping. *Personality and Individual Differences, 54*(7), 865-868.
doi:10.1016/j.paid.2012.12.008
- Triplet, R. G. (1992). Discriminatory Biases in the Perception of Illness: The Application of Availability and Representativeness Heuristics to the AIDS Crisis. *Basic and Applied Social Psychology, 13*(3), 303-322. doi:10.1207/s15324834basp1303_3
- Welch, K. (2007). Black criminal stereotypes and racial profiling. *Journal of Contemporary Criminal Justice, 23*(3), 276-288. doi:10.1177/1043986207306870
- Welsh, M., Burns, N., & Delfabbro, P. (2013). *The cognitive reflection test: How much more than numerical ability?* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.
- Zimmerman, I., Redker, C., & Gibson, B. (2011). The role of faith in intuition, need for cognition and method of attitude formation in implicit–explicit brand attitude relationship strength. *Journal of Consumer Psychology - J CONSUM PSYCHOL, 21*, 290-301.
doi:10.1016/j.jcps.2011.04.001

Tables

Table 1*Descriptive Statistics for Study Measures*

Measure	Mean (SD)
JBT*	0.19 (0.39)
FPST*	0.15 (0.34)
NEDS	3.32 (1.03)
FIS	4.79 (0.99)
Stroop	0.42 (0.31)
Letter Memory	89.96% (7.13)
CRT	40.69% (37.32)
NFC	5.92 (1.05)
SNS	4.10 (0.85)

*Criterion bias difference scores

Table 2*Correlations Between JBT Criterion Bias and Other Study Measures (n=77)*

Measure	Correlation with JBT Criterion Bias (r)	p-value
FPST	.100	.933
NEDS	.179	.119
FIS	.246	.031
Stroop	-.026	.824
Letter Memory	.069	.55
CRT	-.061	.599
NFC	.045	.700
SNS	.213	.062

Appendix A: NED scale items

Participants were asked the following questions:

1. A famous millionaire commissioned a portrait to a local artist for two million dollars. The portrait will be positioned in the hall of a new museum that the millionaire has recently funded, to acknowledge his contribution. The artist portrays him as a triumphant Roman emperor. Based on the information provided, how confident are you that the artist admires the millionaire? (1= Not at All Confident, 7= Very confident)
2. Ben and Zach are presenting introductory information to new employees at a telemarketing company that calls potential customers with information about a home appliance. They flipped a coin to randomly pick what topics each would present. Ben talks about sales strategies that most often do not work. He recounts stories of being hung up on, insulted repeatedly, and led on wild goose chases by people. Zach talks about sales strategies that seem to work well. He relates times that he has connected with people, built good impressions about the product, and arranged large orders with new customers. Based on the available information, how confident are you that Zach sells more than Ben? (1= Not at All Confident, 7= Very confident)
3. Mary is working on an essay on the negative aspects of capitalism that she was assigned to prepare for her sociology course. The main points she plans to include in her essay are that capitalism is inherently exploitative, that it leads to imperialism and oppression, and that it creates wasteful practices such as planned obsolescence of products. Based on the information provided, how confident are you that Mary's attitude toward capitalism is negative? (1= Not at All Confident, 7= Very confident)

4. A struggling freelance writer finally lands her first paid gig. Her employer, a political magazine, assigns her to write a piece advocating for the election of Senator Smith. Her feature story focuses on these three issues: (1) Senator Smith is backing legislation to spur job creation in certain sectors; (2) Senator Smith is committed to reducing America's dependence on foreign oil; and (3) Senator Smith is supporting tax cuts for small businesses. Based on the information provided, how confident are you that the writer supports Senator Smith? (1= Not at All Confident, 7= Very confident)
5. Supermodel Kate Moss has been a celebrity endorser for Rimmel makeup products for over 10 years. She appeared in more than 20 television commercials and in dozens of print advertisements. Based on the information provided, how confident are you that she really likes Rimmel products? (1= Not at All Confident, 7= Very confident)
6. Wendy found herself crying while watching a sentimental movie, which critics raved to be one of the most brilliant, powerful, and emotionally stirring films in cinema history. Based on this information, how confident are you that Wendy is an emotional person? (1= Not at All Confident, 7= Very confident)
7. Deborah had invited Peter to watch her class performance of Shakespeare's play *A Midsummer Night's Dream*. After the show, Deborah and Peter were discussing the plot. During this discussion, Deborah had to correct Peter multiple times on important events that occurred during the play. Based on the information provided, how confident are you that Deborah is more knowledgeable than Peter? (1= Not at All Confident, 7= Very confident)
8. In response to citywide budget cuts a principal is instructed to fire six teachers at her school. The principal fires Mary, a young English teacher who is extremely popular with

the students. Mary is devastated to be unemployed and nervous about how she will pay her rent next month. Based on the available information, how confident are you that the principal is an inconsiderate person? (1= Not at All Confident, 7= Very confident)

9. Lily is in the second grade. Her teacher has given all of the second grade students an assignment to perform an act of charity and tell the class about it. Lily collects \$2.50 in change from her family and neighbours, and donates it to the Salvation Army. Based on the information provided, how confident are you that Lily is a generous child? (1= Not at All Confident, 7= Very confident)
10. Paula and Jasmine live in different suburbs of Los Angeles. Paula's suburb collects trash and recycling separately every week, so Paula takes the time to separate recyclable paper, plastic, and glass from her trash. Jasmine's suburb does not pick up recycling, so she puts all her trash in one trash bag. Based on the information provided, how confident are you that Paula cares about the environment more than Jasmine? (1= Not at All Confident, 7= Very confident)

Appendix B: Faith in Intuition Scale Items

Participants were asked to rate the extent to which they agree with each statement as follows:

1=Strongly disagree

2=Moderately disagree

3=Slightly disagree

4=Neither agree nor disagree;

5=Slightly agree

6=Moderately agree;

7=Strongly agree

The items participants rate are as follows:

I trust my initial feelings about people.

I believe in trusting my hunches.

My initial impressions of people are almost always right.

When it comes to trusting people, I can usually rely on my 'gut feelings'.

I can usually feel when a person is right or wrong, even if I can't explain how.

Appendix C: Statement of Contribution

I assisted in defining measures of interest in this project by reviewing relevant literature. I composed the REB approval for this project, with guidance from Dr. Jordan Axt. I also composed the study pre-registration document, defining how all study measures would be scored and what data analysis would be conducted. Although basic Inquisit 5 code was available on Millisecond for many of the tasks administered, I modified and debugged much of this code. I coded all task instructions, coded several questionnaires, modified task parameters and tested the software extensively both by myself and by running beta participants. I also developed experimenter protocol for running participants and trained other researchers in this protocol. I ran dozens of participants with assistance from lab manager Tony Feng and research assistant Ethan Forgrave. Data analysis was conducted by Dr. Axt, which provided the data in Tables 1 and 2 as well as additional correlations reported in the text. Finally, I composed this paper with extensive comments and feedback from Dr. Axt throughout the process.