Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Full Length Article

Understanding mechanisms behind discrimination using diffusion decision modeling[☆]

Jordan R. Axt[a,*], David J. Johnson[b]

[a] McGill University, Canada
[b] University of Maryland at College Park, United States of America

ARTICLE INFO

Keywords:
Diffusion decision model
Judgment bias task
Intervention
Discrimination

ABSTRACT

Past research has documented where discrimination occurs or tested interventions that reduce discrimination. Less is known about how discriminatory behavior emerges and the mechanisms through which successful interventions work. Two studies ($N > 4500$) apply the Diffusion Decision Model (DDM) to the Judgment Bias Task, a measure of discrimination. In control conditions, participants gave preferential treatment (acceptance to a hypothetical honor society) to physically attractive applicants. DDM analyses revealed participants initially favored attractive candidates and attractiveness was accumulated as evidence of being qualified. Two interventions—raising awareness of bias and asking for more deliberative judgments—reduced discrimination through separate mechanisms. Raising awareness reduced biases in drift rates while increasing deliberation raised decision thresholds. This work offers insight into how discrimination emerges and may aid efforts to develop interventions to lessen discrimination.

Understanding Mechanisms Behind Discrimination Using Diffusion Decision Model Analyses Discrimination, differences in treatment based on social group membership (Axt & Lai, 2019), has far-reaching impacts in workplace and interpersonal contexts. Studies have documented discrimination in many contexts, including gender in student evaluations (Milkman, Akinola, & Chugh, 2015), race in housing accommodations (Turner & Ross, 2003), or sexual orientation in lending decisions (Sun & Gao, 2019).

Lab and field studies have created interventions to reduce discrimination, such as raising awareness of the possibility of bias (Axt, Casola, & Nosek, 2019; Pope, Price, & Wolfers, 2018), providing strategies to alter behavior in the face of biasing information (Mendoza, Gollwitzer, & Amodio, 2010), or committing to decision-relevant criteria beforehand (Uhlmann & Cohen, 2005). Combinations of interventions have also proven effective at reducing the impact of discrimination (Carnes et al., 2012; Devine et al., 2017). While identifying where discrimination occurs and developing interventions to reduce it is valuable, less progress has been made documenting the psychological processes by which social information leads to discrimination and how these processes are disrupted by interventions.

1. Disentangling mechanisms behind discrimination

Consider a hiring manager who, when reviewing candidates to interview, encounters social information before decision-relevant information. This could occur when the information is presented first, such as a name suggesting race at the top of a resume, or prioritized in attention, such as first noticing a headshot indicating gender on a LinkedIn page (Jaeger, Sleegers, Evans, Stel, & van Beest, 2019). How might social information influence selection? One possibility is that social information creates an initial preference for one response. For instance, seeing a physically attractive face might predispose the evaluator to give beneficial treatment, even if they manage to disregard attractiveness for the rest of the decision-making process.

Another not mutually exclusive possibility is that evaluators may (intentionally or unintentionally) use attractiveness in the decision-making process. This could occur in a number of ways, such as treating attractiveness itself as qualification-relevant evidence or through a “halo effect” (Nisbett & Wilson, 1977), where attractiveness changes how more relevant information is interpreted (e.g., making a GPA seem more impressive when the candidate is physically attractive). In both cases, attractiveness information is accumulated throughout judgment

[☆] This paper has been recommended for acceptance by Michael Kraus.

[*] Corresponding author.
E-mail address: jordan.axt@mcgill.ca (J.R. Axt).

https://doi.org/10.1016/j.jesp.2021.104134
Received 1 September 2020; Received in revised form 10 March 2021; Accepted 15 March 2021

in addition to the candidate's actual qualifications, resulting in beneficial treatment for more attractive applicants.

Though both accounts—an initial preference for more attractive people versus attractiveness being used in the decision-making process—result in discrimination, they likely require different interventions to change behavior.

The present studies examine these biases within the Judgment Bias Task (JBT; (Axt, Nguyen, & Nosek, 2018). In the JBT, participants evaluate a series of profiles, here applicants for an academic honor society. Applicants are presented with qualification-relevant criteria (e.g., GPA) and irrelevant social information. Qualifications are manipulated so some applicants are more deserving than others. Since the JBT has objectively correct and incorrect decisions, researchers can evaluate how social information impairs selection of more deserving applicants. Prior studies using the JBT find discrimination based on ingroup status or physical attractiveness (Axt et al., 2019), mirroring biases in field settings (King & Ahmad, 2010; Rooth, 2009).

2. Signal detection model

Previous work has revealed two distinct approaches for reducing discrimination within the JBT. Using Signal Detection Theory (SDT), performance on the JBT has been modeled in terms of *bias* and *noise*.

Bias occurs when one social group is more likely to get a favorable response (e.g., acceptance to an honor society) than another social group. SDT analyses model bias through the criterion parameter, with relatively lower values indicating a lower bar for giving a favorable response to members of a social group. Because the response criterion is on average lower for more versus less physically attractive applicants (Axt et al., 2018), physically attractive applicants are relatively more likely to be accepted when not qualified and less physically attractive applicants are relatively more likely to be rejected when qualified.

In contrast, noise refers to how well evaluators accurately distinguish between more versus less qualified applicants. SDT analyses model noise in decision-making through the sensitivity parameter, with greater noise meaning lower sensitivity.

Within the SDT approach, the magnitude of discrimination in the JBT is dependent on the degree of sensitivity (how many applicants are incorrectly accepted or rejected) and differences in criterion (the degree to which those incorrect decisions favor one social group). Prior studies (Axt & Lai, 2019) have found various interventions differently impact these outcomes. For instance, participants made aware of the tendency to favor more attractive people showed reduced bias in criterion but no change in sensitivity on a subsequent JBT. That is, raising awareness about favoritism based on attractiveness did not reduce judgment errors, it only made those errors more fairly distributed.

Conversely, participants who were required to slow down (or speed up) their judgments showed higher (or lower) levels of sensitivity but no changes in criterion biases. Similar results emerged when participants were told to think more deliberatively about their decisions. These timing and deliberation interventions impacted the total number of errors made in judgment, but those remaining errors still favored attractive applicants at the same rate as among participants making judgments at their own pace.

While these SDT analyses are informative, they do not speak to the dynamic process behind how discrimination emerges and can be reduced throughout the course of the decision. That is, an SDT framework is not well equipped to tease apart whether discrimination arises from a process of initial preferences for certain candidates, relying on social information as evidence, or both.

This limitation is because SDT is a static analysis that ignores information about the length of the decision process. Here, we use a dynamic decision model to disentangle how social information leads to discrimination in judgment and how various interventions impact this process.

3. Diffusion decision model

The Diffusion Decision Model (DDM; Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016) is a sequential sampling model used to explain the process underlying decisions in two-choice tasks by simultaneously modeling choices and their speed. The DDM decomposes decisions into four components: relative start point (β), threshold separation (α), drift rate (δ), and non-decision time (τ). See Table 1 for a description of model parameters and Fig. 1 for a graphic summary.

In the JBT, faces and qualifications appear simultaneously. If face perception precedes encoding of qualifications (Farah, Wilson, Drain, & Tanaka, 1998; Tsao & Livingstone, 2008), then social information communicated through a face could shift participants' relative start point (β) to initially favor acceptance or rejection. Participants then repeatedly sample the stimulus for relevant evidence, which could reflect qualifications but also be influenced by social information like physical attractiveness. The average rate of evidence accumulation is reflected by the drift rate (δ), with positive (negative) values indicating evidence to accept (reject). Evidence is accumulated until participants hit either the accept or reject threshold boundary (α), at which point they render the corresponding choice. Finally, the proportion of the minimum response time that is unrelated to decision-making (e.g., motor response time) is indicated by the non-decision time parameter (τ').

An advantage of the DDM is that it can test process-level accounts of behavior. This is useful when competing accounts predict similar behavior for different reasons. For example, research on shooting/weapon identification tasks has shown social information (race) causes participants to misidentify harmless objects as weapons. When race is presented prior to the critical object, it impacts the relative start point (Todd et al., in press). Conversely, when race is presented simultaneously with the critical object, it typically impacts evidence accumulation (Correll, Wittenbrink, Crawford, & Sadler, 2015; Pleskac, Cesario, & Johnson, 2018).

However, similar analyses have not been completed on more controlled and ambiguous judgments, which may better reflect how discrimination occurs in many domains. Consider again the hiring manager using LinkedIn to evaluate applicants. Discrimination could occur because of an initial preference for physically attractive candidates and/or because attractiveness is incorporated into decision-making. The latter could mean that attractiveness is directly treated as a "hidden" qualification directly weighted alongside more relevant evidence like education, or that attractiveness could indirectly make existing qualifications seem more positive.

Better understanding processes underlying discriminatory behavior brings theoretical and practical benefits. For instance, the most prominent model of socially biased judgment is Wilson and Brekke's (1994) work on "mental contamination." According to this perspective, biased judgments are initiated when unwanted mental processing is triggered. However, the model is ambiguous as to when unwanted mental processing occurs during the decision process. DDM analyses can refine this model by revealing when such unwanted mental processing occurs:

Table 1
Parameters of the diffusion decision model (DDM).

Parameter	Interpretation
Threshold separation (α)	The separation between the two thresholds, determining the amount of evidence required to decide, with $0 < \alpha$.
Relative start point (β)	Initial bias to accept candidates at the start of the evidence accumulation process, with $0 < \beta < 1$. Values above 0.50 indicate a bias to accept.
Drift rate (δ)	Average quality of evidence extracted from a stimulus at each unit of time, with $\infty < \delta < \infty$. Higher absolute values indicate stronger evidence. Positive values indicate evidence to accept.
Non-decision time (τ')	Proportion of the minimum response time spent on processes un- related to decision-making, with $0 < \tau^t < 1$.

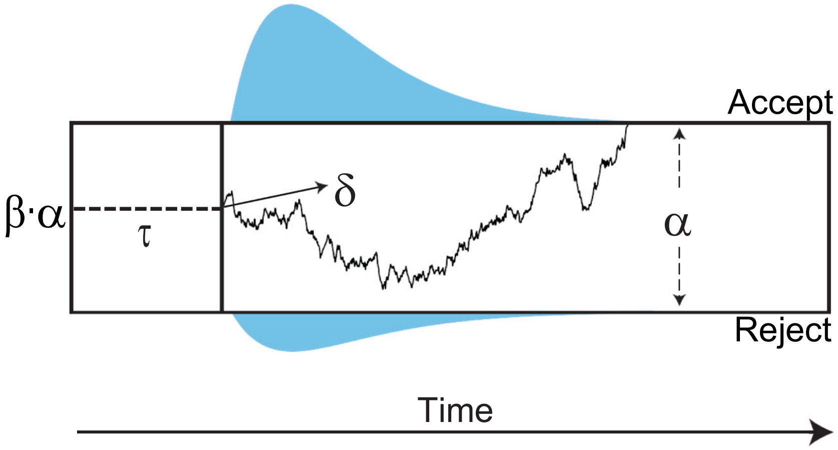

Fig. 1. The decision diffusion model as applied to the JBT. Individuals start with an initial bias β to accept the candidate or not. Noisy information is accumulated over time with average strength δ . The amount of information needed to make a decision is indicated by the threshold separation α. The length of non-decision processes is indicated by τ'. The model predicts the distribution of response times for accept and reject decisions (in blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

immediately when social information is first encountered, over time as evidence is accumulated, or both.

Similarly, the way in which the DDM divides the decision process can reveal connections to other perspectives in the decision-making literature. For instance, if discrimination on the JBT maps on to differences in the relative start point but not in evidence accumulation (i.e., drift rates), it would be consistent with an anchoring and adjustment account (Epley & Gilovich, 2006). In this account, the impact of biasing information occurs immediately, with discrimination emerging when people fail to sufficiently adjust for these initial preferences. A lack of differences in drift rate would be consistent with attractiveness not influencing judgment beyond initial preferences. Alternatively, if differences in drift rates based on attractiveness exist under control conditions but are reduced following certain interventions, it would suggest that the intervention promotes self-regulatory processes that provide participants with the ability to counteract their own prejudiced responses (e.g., Monteith & Mark, 2005).

Practically, the DDM can clarify how interventions to reduce discrimination work. It is currently unclear how prominent bias-reducing interventions disrupt psychological processes leading to discrimination. For example, heightened awareness could reduce an initial preference for certain group members (impacting the relative start point), reduce the degree to which social information is used as information during evidence accumulation (impacting the drift rate), or both. Similarly, interventions designed to reduce discrimination by minimizing errors (e.g., by delaying responses or inducing more deliberate decision-making; Axt & Lai, 2019) may have distinct effects on the decision processes. For instance, these interventions could shift evaluators to focus on accuracy over speed (impacting threshold separation).

Here, we conduct DDM analyses on archival and novel JBT data. Specifically, participants evaluated applicants for an academic honor society where candidates varied in physical attractiveness. We analyze control conditions and a series of interventions that have been shown to impact the magnitude of discrimination on the JBT (Axt et al., 2019; Axt & Lai, 2019).

4. Analytic approach

Although not the focus of our hypotheses, we first present analyses of acceptance decisions in order to put the results of the SDT and DDM models in context. Decisions were analyzed with multilevel logistic regression (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2020).

Models include random intercepts for participants and targets, and random slopes by participant for attractiveness and qualification (Barr, Levy, Scheepers, & Tily, 2013; Judd, Westfall, & Kenny, 2012). Interventions were dummy coded and condition means transformed to proportions to reflect the likelihood of accepting an applicant.

To allow for comparisons to prior work, we briefly report SDT analyses in addition to DDM analyses. SDT analyses examine decisions only (not response times), and followed the same procedure as prior work (Axt & Lai, 2019), although trials over 15,000 milliseconds were removed to be consistent with the data used in the DDM analyses.

Decision and response time data were simultaneously analyzed with a multilevel DDM using Bayesian methods (Johnson, Hopwood, Cesario, & Pleskac, 2017; Plummer, 2003; Vandekerckhove, Tuerlinckx, & Lee, 2011; Wabersich & Vandekerckhove, 2014). We allowed all DDM parameters to vary by intervention and candidate attractiveness. A model comparison approach using the Deviance Information Criterion (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) demonstrated that a model where the drift rate alone changed as a function of qualifications best fit the data.

We describe the posterior distribution of a parameter by reporting the most credible value and 95% Highest Density Interval (HDI). When testing condition differences and interactions we report the most credible estimate, the effect transformed to a standardized estimate *d*, and its 95% HDI. Effect sizes were calculated by dividing differences by the group-level subject variability estimate for that parameter.

Parameters were given uninformative priors. We did not estimate across-trial parameter variability because participants saw each stimulus only once and the small number of trials per condition (16) prevented more complex models from converging. Posterior predictive model fits predicted acceptance rates and response time distributions well although the model underestimates error response times. We believe these misses are due to small sample sizes: each condition had only 16 trials, of which approximately 30% were errors. Model fits, regression tables, code, diagnostics, and parameter estimates are provided in the Supplemental Materials.

5. Study 1

5.1. Methods

We reanalyzed three studies from Axt and Lai (2019) that tested interventions to reduce discrimination in the JBT. We included the timed (Study 2b), deliberative (Study 4), and awareness (Study 5) interventions. Each study included a control condition for comparison. These interventions were chosen given the prior evidence that each has a distinct (and even opposing) effect on the decisions made in the JBT (2019), with the timed and deliberative interventions impacting the number of errors made but not the distribution of those errors and the awareness intervention impacting the distribution of errors but not the number of errors made.

5.1.1. Participants

Participants ($N = 3681$) came from the Project Implicit research pool (Nosek, 2005). The archival design of Study 1 did not allow us to specify a target sample size. We removed 341 participants who accepted less than 20% or more than 80% of candidates (Axt & Lai, 2019). We also removed 1.0% trials where participants responded under 300 ms or over 15 s (1.8 s in the timed intervention). Finally, we removed 37 participants with less than 50% of trials after exclusions. The final sample was 3303 participants ($N_{\text{Speed}} = 594$; $N_{\text{Deliberative}} = 478$; $N_{\text{Awareness}} = 539$; $N_{\text{Control}} = 1692$). Participants were 65% women, 70% White, with a mean age of 33 ($SD = 14$). This sample size provided at least 80% power for detecting a between-subjects effect as small as $d = 0.15$ when comparing each intervention to the control condition, and at least 80% power for detecting any within-subjects effect as small as $d = 0.13$.

5.1.2. Procedure

Study procedures can be found in (Axt & Lai, 2019). Aside from the JBT, participants completed measures of perceived and desired JBT performance, as well as measures of explicit and implicit attractiveness associations. For all studies, data for these measures are available at https://osf.io/psazg but are not included in primary analyses. For both studies, we report all measures, manipulations and exclusions. We analyzed data from all participants that finished the JBT.

In the JBT, participants were told to accept half of 64 hypothetical academic honor society applicants. Each applicant had four pieces of information (Science GPA, humanities GPA, interview score, and recommendation letter strength). Participants were told to weigh each piece of information equally. Half of the applicants were scored to be more qualified and half less qualified. Each applicant had a unique combination of qualifications.

Applicants were shown with a headshot depicting White, smiling, college-aged people. These photos had been pre-tested to vary in physical attractiveness (Axt et al., 2018). For both more and less qualified applications, there were an equal number of more and less physically attractive applicants, which were also evenly divided between males and females.

Participants first viewed each application for one second during an encoding phase. Next, participants saw each application one at a time and completed an untimed judgment (except for in the timed condition) to accept or reject the applicant. Each application was equally likely to be paired with a more or less physically attractive face across 12 presentation orders.

5.1.3. Interventions

We tested three interventions. In the timed intervention, participants were asked to make decisions quickly and trials timed out after 1800 ms. Participants in the deliberative intervention were told to “think hard and slow down” when making their evaluations, though no time delays were imposed. Participants in the awareness intervention were warned about biases favoring more attractive applicants and asked to avoid letting physical attractiveness impact their decisions. Participants in control conditions completed the JBT without additional instructions.

5.2. Results

5.2.1. Behavioral results

Qualified applicants were more likely to be accepted ($M = 0.68$ [0.67, 0.70]) than unqualified applicants ($M = 0.34$ [0.32, 0.36]), $b =$ 1.64 [1.60, 1.67]. Relative to the control condition, participants were more accurate (i.e., more likely to accept qualified candidates) in the deliberate intervention, $b = 0.15$ [0.07, 0.23], and less accurate in the timed intervention, $b = -0.62$ [−0.69, −0.55].

In the control condition, attractive applicants were more likely to be accepted ($M = 0.54$ [0.52, 0.55]) than unattractive applicants ($M = 0.49$ [0.47, 0.51]), $b = 0.21$, [0.10, 0.32]. The effect of attractiveness on judgments was reduced in the awareness intervention, $b = -0.21$ [−0.28, −0.13], where attractive applicants were not more likely to be accepted ($M = 0.51$ [0.49, 0.53]) than unattractive applicants ($M = 0.51$ [0.49, 0.53]), $b = 0.01$ [−0.10, 0.12]. The effect of attractiveness did not vary in the deliberate intervention $b = -0.06$ [−0.13, 0.02], nor in the timed intervention $b = 0.05$, [−0.02, 0.12].

5.2.2. SDT results

We briefly report SDT analyses to allow comparisons to prior work Axt and Lai (2019); see the online supplement for full analyses. Replicating past studies, there were no criterion differences based on candidate attractiveness in the awareness intervention ($d = 0.02$); in all other conditions the criterion for more physically attractive applicants was lower than the criterion for less physically attractive applicants (all d's > 0.24). Relative to the control condition, only the awareness condition showed a reliable reduction in criterion biases favoring more over less attractive applicants ($d = 0.28$), while only the timed condition decreased overall sensitivity ($d = -0.79$) and only the deliberate condition increased overall sensitivity ($d = 0.19$).

5.3. DDM results

5.3.1. Attractiveness and qualifications

Fig. 2 presents the four DDM parameters for each condition and level of applicant attractiveness. Collapsing across conditions, the threshold separation did not vary by candidate attractiveness (top left panel), $b =$ −0.01, $d = -0.01$ [−0.03, 0.01], nor did the proportion of the minimum response time spent on non-decision processes (bottom left panel) $b =$ −0.001 [−0.006, 0.004]. In contrast, the relative start point (top right panel) was slightly higher for more attractive ($M = 0.493$, [0.489, 0.496]) versus less attractive candidates ($M = 0.483$, [0.480, 0.487]), b = 0.010, $d = 0.14$ [0.09, 0.18], an effect that reflects an initial preference to choose more over less attractive candidates.

The drift rate (Fig. 2, bottom right panel) was also higher for more attractive ($M = 0.07$, [0.06, 0.08]) versus less attractive candidates (M = 0.02, [0.01, 0.03]), $b = 0.05$, $d = 0.58$ [0.44, 0.71]. This finding is consistent either with attractiveness being treated as a form of qualification-relevant evidence or a halo effect where more attractive candidates' qualifications are perceived as more impressive. Finally, the drift rate was much higher for qualified candidates ($M = 0.31$, [0.31, 0.32]) versus unqualified candidates ($M = -0.22$, [−0.23, −0.22]), $b =$ 0.54, $d = 6.63$ [6.28, 7.07].

5.3.2. Timed intervention

The timed intervention impacted each of the DDM parameters. The threshold separation in the timed intervention ($M = 1.54$ [1.48, 1.59]) was smaller than the control condition ($M = 3.01$ [2.97, 3.05]), $b =$ −1.47, $d = -1.85$ [−1.96, −1.75], and the proportion of time spent on non-decision processes in the timed intervention $M = 0.891$ [0.889, 0.894]) was larger than the control condition ($M = 0.720$ [0.716, 0.724]), $b = 0.172$ [0.167, 0.176]). In other words, non-decision processes took up a larger *proportion* of response times when judgments were shortened by a response window. Drift rates were stronger (i.e., farther from zero) in the timed intervention $b = 0.070$, $d = 0.89$ [0.0.67, 1.09], and the start point was credibly higher in the timed intervention ($M = 0.500$ [0.493, 0.506]) than in the control condition ($M = 0.482$ [0.478, 0.486]) $b = 0.017$, $d = 0.25$ [0.14, 0.35].

The effect of attractiveness on the drift rate was stronger in the timed intervention versus the control condition; specifically, evidence to select more attractive candidates accumulated more quickly ($b = 0.116$, $d =$ 1.45 [0.98, 1.87]). The effect of attractiveness on the relative start point was weaker in the timed intervention, $b = -0.014$, $d = -0.20$ [−0.31, −0.09], although this finding did not replicate in Study 2.

5.3.3. Deliberate intervention

The deliberate intervention generally had the opposite impact of the timed intervention. Relative to the control condition, the threshold

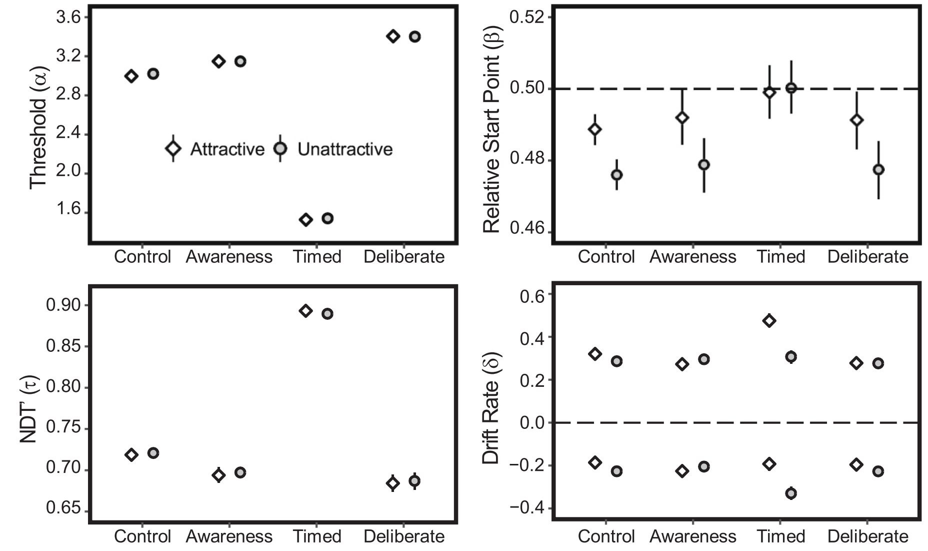

Fig. 2. Effect of attractiveness and condition on Study 1 threshold (top left), relative start point (top right), proportion non-decision time (NDT'; bottom left), and drift rate (bottom right). Drift rates for unqualified candidates appear below drift rates for qualified candidates. Points are predicted means and bars are 95% HDI.

separation in the deliberate intervention was larger (M = 3.40 [3.34, 3.47]), b = 0.39, d = 0.51 [0.40, 0.61]). The proportion of time spent on non-decision processes in the deliberate intervention was smaller (M = 0.686 [0.679, 0.693]), b = −0.034 [−0.042, −0.026]). However, the overall strength of drift rates did not vary in the deliberate intervention from the control condition, b = −0.011, d = −0.14 [−0.27, 0.02], nor did the start point credibly vary from the control condition b = 0.002, d = 0.03 [−0.09, 0.14].

The deliberate intervention only moderated the effects of attractiveness on the drift rate, b = −0.02, d = −0.29 [−0.51, −0.05]. Evidence to select more attractive candidates accumulated less quickly in the deliberate intervention, although this finding did not replicate in Study 2.

5.3.4. *Awareness intervention*

The awareness intervention had similar impacts as the deliberate intervention. Relative to the control condition, the threshold separation in the awareness intervention was larger (M = 3.15 [3.08, 3.21]), b = 0.14, d = 0.17 [0.07, 0.27]), although this effect did not replicate in Study 2. The proportion of time spent on non-decision processes was smaller (M = 0.720 [0.716, 0.724]), b = −0.024 [−0.032, −0.017]). The overall strength of drift rates did not vary in the deliberate intervention from the control condition, b = −0.006, d = −0.07 [−0.21, 0.08], nor did the start point credibly vary from the control condition b = 0.002, d = 0.03 [−0.07, 0.16].

The awareness intervention only moderated the effects of attractiveness on the drift rate, b = −0.06, d = −0.73 [−0.96, −0.49], such that evidence to select more attractive candidates accumulated less quickly in the awareness intervention.

5.4. *Discussion*

In the control condition, participants were more likely to accept attractive students. The awareness intervention reduced the bias to accept more attractive students but did not increase accuracy. In contrast, the deliberate intervention increased accuracy and the timed intervention reduced it, though neither intervention impacted bias.

DDM analyses revealed that in the control condition, the evidence accumulation rate towards acceptance was stronger for more versus less attractive candidates. There was also a small bias to initially favor the accept decision for more versus less attractive candidates. The decrease in bias to accept more attractive students in the awareness intervention was reflected by attractiveness having less of an impact on the evidence accumulation process (i.e., the drift rate parameter).

Accuracy changes in the deliberate and timed interventions were reflected in the threshold separation and drift rate parameters. Participants' threshold separations were more than twice as large in the deliberate intervention than the timed intervention, although drift rates were also far stronger (d = 0.89) in the timed intervention than in the control condition. These results reflect why error rates—although higher in the timed intervention than the control or deliberation intervention—were not substantially higher, as participants collected evidence at a faster rate when they spent less time making decisions. Similar relationships between DDM parameters as a function of response window are well-documented in the decision-making literature (Rae, Heathcote, Donkin, Averell, & Brown, 2014; Vandekerckhove et al., 2011).

The timed intervention also impacted the influence of attractiveness on the evidence accumulation process. Relative to the control condition, attractiveness had a larger effect on the drift rate in the timed intervention. However, while the timed intervention impacted the effect of attractiveness on the drift rate, it did not reduce the impact of attractiveness on actual decisions. That is, the effect of the timed intervention on the evidence accumulation process relative to the control condition (d = 1.45) may not have translated into changes in preferences for more attractive applicants because of the substantially smaller threshold separation that was also created by the timed manipulation (d = −1.85).

In the DDM, decisions are a function of both the evidence accumulation *rate* and evidence accumulation *duration* (i.e., drift rate and threshold separation). When the duration of evidence accumulation is the same, higher drift rates to accept attractive candidates would lead to a higher likelihood of accepting more versus less attractive applicants. However, the impact of these higher drift rates can be offset when the duration of evidence accumulation is shorter. In other words, while the timed intervention may have strengthened the degree to which attractiveness was used during the decision-making process, the shorter response window also reduced the amount of time that such a bias could operate.

One limitation of these results is that they are not a true experiment. Each intervention was compared against several pooled control conditions, which may obscure possible history effects. We addressed these issues in Study 2 by randomly assigning participants to condition.

6. Study 2

6.1. Methods

We replicated Study 1 using a randomized experiment. This study was pre-registered, using $\alpha = 0.05$ in all analyses: https://osf.io/bwnj2.

6.1.1. Participants

Participants ($N = 1558$) were from the Project Implicit research pool. Sample size was determined before any data analysis. We removed 149 participants who accepted less than 20% or more than 80% of candidates (Axt & Lai, 2019). To reduce careless responses we removed 0.4% trials where participants responded under 300 ms or over 15 s (1.8 s in the timed intervention). No participant had less than 50% of trials after these exclusions and so all participants' data were included. The final sample was 1408 participants ($N_{\text{Speed}} = 379$; $N_{\text{Deliberative}} = 329$; $N_{\text{Awareness}} = 360$; $N_{\text{Control}} = 341$). Participants were 65% women, 70% White, with a mean age of 32 ($SD = 14$). This sample size provided at least 80% power for detecting a between-subjects effect as small as $d = 0.22$ when comparing each intervention to the control condition, and at least 80% power for detecting within-subjects effect as small as $d = 0.15$.

6.1.2. Procedure

The procedure was identical to Study 1, except participants were randomly assigned to condition.

6.2. Results

6.2.1. Behavioral results

Qualified applicants were more likely to be accepted ($M = 0.64$ [0.62, 0.66]) than unqualified applicants ($M = 0.34$ [0.32, 0.36]), $b = 1.59$ [1.51, 1.67]. Participants were less likely to accept applicants in the timed intervention ($M = 0.49$ [0.47, 0.50]) than in the control condition ($M = 0.52$ [0.50, 0.53]), $b = -0.14$ [−0.23, −0.06]. Relative to the control condition, participants were more accurate (i.e., more likely to accept qualified candidates) in the deliberate intervention, $b = 0.15$ [0.03, 0.27], and less accurate in the timed intervention, $b = -0.66$ [−0.77, −0.55].

In the control condition, more attractive applicants were more likely to be accepted ($M = 0.55$ [0.53, 0.57]) than less attractive applicants ($M = 0.49$ [0.47, 0.51]), $b = 0.28$, [0.15, 0.41]. The effect of attractiveness was reduced—but not eliminated—in the awareness intervention, $b = -0.14$ [−0.25, −0.03], where attractive applicants were slightly more likely to be accepted ($M = 0.53$ [0.50, 0.55]) than unattractive applicants ($M = 0.50$ [0.47, 0.52]), $b = 0.14$ [0.02, 0.27]. Relative to the control condition, the effect of attractiveness did not vary in the deliberate intervention $b = -0.02$ [−0.14, 0.09], nor in the timed intervention $b = -0.03$, [−0.14, 0.08].

6.2.2. SDT results

The criterion for more physically attractive applicants was lower than the criterion for more physically attractive applicants across all conditions (see online supplement for full results). Relative to the control condition, only the awareness intervention showed a reliable reduction in criterion biases favoring more over less attractive applicants ($d = 0.20$). In addition, relative to the control condition, the timed intervention decreased overall sensitivity ($d = -0.79$), and the deliberate intervention increased overall sensitivity ($d = 0.17$). These results replicate the findings of both Study 1 and Axt and Lai (2019).

6.3. DDM results

6.3.1. Attractiveness and qualifications

Fig. 3 presents the four DDM parameters for each condition and level of applicant attractiveness. The threshold separation did not vary by candidate attractiveness (top left panel), $b = 0.01$, $d = 0.01$ [−0.02, 0.03], nor did the proportion of the minimum response time spent on non-decision processes (bottom left panel) $b = -0.003$ [−0.03, 0.01]. In contrast, the relative start point (top right panel) was slightly higher for more attractive ($M = 0.489$, [0.484, 0.493]) versus less attractive candidates ($M = 0.481$, [0.477, 0.486]), $b = 0.007$, $d = 0.11$ [0.04, 0.17]. As in Study 1, this result reflects an initial preference to choose more over less attractive candidates.

The drift rate (Fig. 3, bottom right panel) was also higher for more attractive ($M = 0.07$, [0.06, 0.09]) versus less attractive candidates ($M = 0.02$, [0.01, 0.03]), $b = 0.07$, $d = 0.75$ [0.60, 0.91]. Replicating Study 1, this finding is consistent with attractiveness being treated as either qualification-relevant evidence or through a halo effect where more attractive candidates' qualifications are seen as more impressive. Finally, the drift rate was much higher for qualified candidates ($M = 0.33$, [0.32, 0.34]) versus unqualified candidates ($M = -0.22$, [−0.23, −0.21]), $b = 0.55$, $d = 5.58$ [5.13, 6.10].

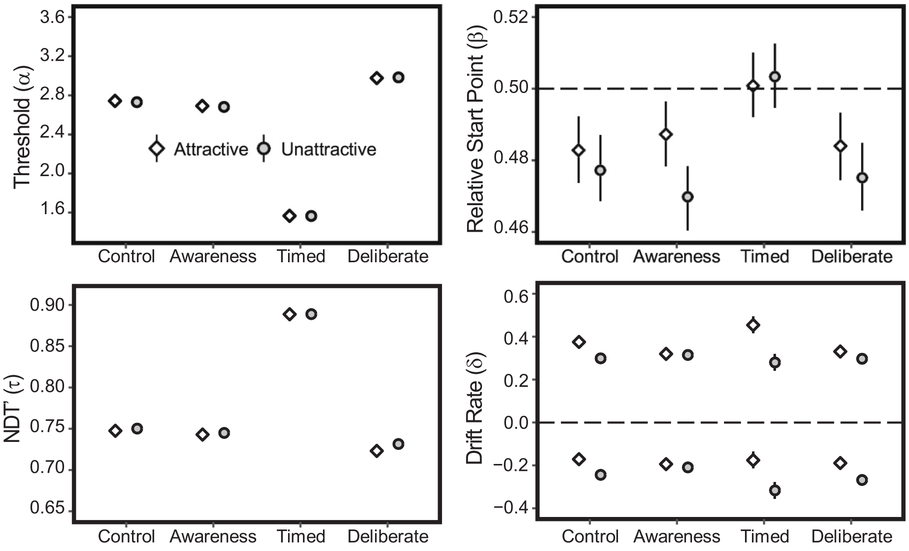

Fig. 3. Effect of attractiveness and condition on Study 2 threshold (top left), relative start point (top right), proportion non-decision time (NDT'; bottom left), and drift rate (bottom right). Drift rates for unqualified candidates appear below drift rates for qualified candidates. Points are predicted means and bars are 95% HDI.

6.3.2. Timed intervention

The timed intervention impacted each of the DDM parameters. The threshold separation in the timed intervention (M = 1.57 [1.51, 1.62]) was smaller than the control condition (M = 2.74 [2.67, 2.80]), b = −1.17, d = −1.95 [−2.11, −1.77]. The proportion of time spent on non-decision processes in the timed intervention (M = 0.889 [0.886, 0.892]) was larger than the control condition (M = 0.749 [0.742, 0.756]), b = 0.140 d = 0.55, [0.52, 0.58]). Drift rates were overall stronger (i.e., farther from zero) in the timed intervention b = 0.036, d = 0.33 [0.0.11, 0.61], and the start point was credibly higher (M = 0.502 [0.494, 0.510]) than in the control condition (M =0.481 [0.472, 0.488]), b = 0.022, d = 0.33 [0.16, 0.50].

The effect of attractiveness on the drift rate was stronger in the timed intervention versus the control condition. Evidence to select more attractive candidates accumulated more quickly (b = 0.084, d = 0.82 [0.37, 1.33]). Unlike Study 1, the effect of attractiveness on the relative start point was not credibly weaker in the timed intervention, b = −0.007, d = −0.11 [−0.29, 0.07].

6.3.3. Deliberate intervention

Again, the deliberate intervention generally had the opposite impact of the timed intervention. Relative to the control condition, the threshold separation in the deliberate intervention was larger (M = 2.98 [2.91, 3.05]), b = 0.24, d = 0.39 [0.24, 0.55]), and the proportion of time spent on non-decision processes was smaller (M = 0.727 [0.719, 0.735]), b = −0.021, d = −0.07 [−0.10, −0.04]). The overall strength of drift rates did not vary in the deliberate intervention from the control condition, b = 0.000, d = 0.00 [−0.21, 0.21], nor did the start point credibly vary from the control condition b = 0.000, d = 0.00 [−0.18, 0.17]. Finally, unlike Study 1, the deliberate intervention did not credibly reduce the effects of attractiveness on the drift rate, b = −0.023, d = −0.19 [−0.52, 0.13].

6.3.4. Awareness intervention

Unlike Study 1, the threshold separation was not credibly different between the awareness intervention (M = 2.68 [2.62, 2.75]) and the control condition, b = −0.05, d = −0.07 [−0.23, 0.07]). However, the proportion of time spent on non-decision processes was again smaller (M = 0.720 [0.716, 0.724]), b = −0.024 [−0.032, −0.017]). The overall strength of drift rates did not vary in the deliberate intervention from the control condition, b = −0.013, d = −0.13 [−0.35, 0.08]. The start point did not credibly vary from the control condition b = −0.001, d = −0.02 [−0.20, 0.14]. Unlike Study 1, the awareness intervention actually increased the pro-attractive bias in the relative start point, b = 0.013, d = 0.19, [0.02, 0.38]).

Consistent with Study 1, the awareness intervention moderated the effects of attractiveness on the drift rate, b = −0.065, d = −0.65 [−0.99, −0.32] as evidence to select more attractive candidates accumulated less quickly in the awareness intervention.

6.4. Discussion

Replicating Study 1, participants were more likely to accept attractive students. The awareness intervention reduced bias to accept more attractive students but did not increase accuracy. The deliberate intervention increased accuracy and the timed intervention reduced accuracy, though neither intervention impacted bias.

DDM analyses were also largely consistent with Study 1. In the control condition, the evidence accumulation rate to accept was stronger for more versus less attractive candidates. The decrease in bias to preferentially accept more attractive applicants in the awareness intervention was reflected by attractiveness having less of an impact on the evidence accumulation process (i.e., the drift rate parameter). In addition, greater (reduced) accuracy following the deliberate (timed) intervention was reflected in the larger (smaller) threshold separation parameter.

Consistent with Study 1, the timed intervention also impacted the effect of attractiveness on the evidence accumulation process. Relative to the control condition, the timed intervention had a larger impact on the role of attractiveness in the evidence accumulation process (d = 0.82), though the influence of this pro-attractive effect on decisions was again mitigated by a substantially smaller threshold separation parameter (d = −1.95), a combination of results that likely explains why the timed intervention did not lead to an increased attractiveness bias in observed choices.

Finally, in both studies we replicated an overall small bias in the relative start point showing an initial preference for attractive candidates. However, the interventions had inconsistent effects on this bias. In Study 1 the timed intervention reduced this bias, whereas in Study 2 the awareness intervention increased it. Because neither of these effects replicated across studies, we refrain from discussing them further.

7. General discussion

The current work examined mechanisms through which socially biased judgments emerge and are reduced in the JBT. More physically attractive candidates were more likely to be accepted than less physically attractive candidates in control conditions. This discrimination was reflected both in an initial preference for physically attractive applicants (a relative start point effect) and attractiveness being incorporated into decisions (a drift rate effect). The latter is consistent with two explanations: attractiveness is treated as a hidden qualification, or attractiveness positively biases the interpretation of other qualifications similar to a halo effect.

Relative to the control condition, the awareness intervention consistently reduced the impact of attractiveness on the drift rate (Study 1 d = −0.73, Study 2 d = −0.65) while the deliberate intervention increased the overall decision threshold (Study 1 d = 0.51, Study 2 d = 0.39). In contrast, the timed intervention both increased the effect of attractiveness on the drift rate (Study 1 d = 1.45, Study 2 d = 0.82) while also reducing decision threshold due to the addition of a response window (Study 1 d = −1.85, Study 2 d = −1.95).

This pattern of results highlights nuance obscured by analyses using a static approach like SDT. Prior SDT analysis (Axt & Lai, 2019) found the deliberation and timed interventions had similar (if opposing) outcomes of higher or lower sensitivity but no changes on criterion biases. In contrast, the DDM analyses used in this work found the deliberate intervention increased decision thresholds and a timed intervention reduced them, but only the timed intervention also increased the influence of attractiveness on the drift rate. In other words, solely focusing on SDT outcomes between deliberation and timed interventions may gloss over how each manipulation differently attenuates or exacerbates the degree of discrimination found on the JBT.

DDM analyses further revealed complex effects of prioritizing speed in social judgment. Limiting decision-making time increased the amount of errors (by lowering the decision threshold) and heightened reliance on social information during the course of decision-making (by increasing the impact of attractiveness on drift rates). This pattern suggests that speeding the decision process not only leads people to favor speed over accuracy more, as indicated by the threshold parameter, it also makes the social information more influential during the course of decision-making, as indicated by a greater impact of attractiveness on the drift rate.

There were some inconsistencies in the results found in Studies 1 and 2. For example, the deliberate intervention reduced the impact of attractiveness on drift rates in Study 1 but not in Study 2, whereas the awareness intervention raised threshold separation in Study 1 but not Study 2. Considering the identical procedures and sample sources across studies, it is difficult to conclude whether these inconsistencies reflect a false positive (or a false negative) versus some unaccounted variable that explains these divergent findings. Nonetheless, the strongest effects in each study clearly replicated, and we focus our interpretation on effects present in both analyses.

7.1. Implications for discrimination

These analyses may inform efforts to develop discrimination-reducing interventions. Prior JBT studies used signal detection analyses, which do not incorporate response times (Axt et al., 2019) and were thus ambiguous concerning *when* social information shapes judgment. Our analyses show not only does attractiveness lead to an initial preference to accept candidates, participants accumulate such information throughout judgment.

This finding has implications for prominent models of bias. The "mental contamination" framework by Wilson and Brekke (1994) argues biased judgment occurs when unwanted mental processing is initially triggered. Such a claim is consistent with the relative start point bias found here. At the same time, the impact of attractiveness on the drift rate suggests the influence of social information is not limited to the start of decision-making; rather, it impacts evidence collected throughout the decision process. Biased judgment may then not only be the result of an early, faulty mental process, but rather a combination of processes that promote favoritism towards certain groups both initially *and* throughout the judgment process.

Likewise, these results suggest that bias favoring more physically attractive people on the JBT is not merely an example of anchoring and adjustment (Epley & Gilovich, 2006), as the drift rate findings indicate attractiveness continues to exert an influence on judgment past any initial preferences. As a result, one productive avenue for reducing discrimination may be to develop sustained abilities to counteract prejudiced responses (e.g., Monteith, 1993). Future interventions may find greater success by providing frequent reminders to avoid biasing information (e.g., Forscher, Mitamura, Dix, Cox, & Devine, 2017), compared to the one-off interventions used in this work. For example, continued reliance on biasing information may explain why many interventions have difficulties shifting subsequent behavior (e.g., Chang et al., 2019).

More broadly, DDM analyses offer a productive avenue for researchers to more directly test assumptions about mechanisms underlying discriminatory behavior. The structure of the JBT allows slower judgments than those in prior DDM investigations (Pleskac et al., 2018; Todd et al., in press). Despite large response time differences between the timed versus untimed JBT, this manipulation impacted the threshold separation and non-decision time in expected ways, as these parameters are sensitive to response window changes (Pleskac et al., 2018). Such results suggest the DDM could be applied to other areas, such as moral judgments (Brannon, Carr, Jin, Josephs, & Gawronski, 2019) or face perception (Jaeger et al., 2019).

7.2. Reducing the relative start point bias

One notable finding is the small relative start point bias reflecting an initial preference towards accepting more attractive applicants. While awareness and deliberate interventions reduced discrimination by decreasing the impact of attractiveness in evidence accumulation or increasing threshold separation, respectively, they did not consistently reduce the relative start point bias. This start point bias reflects an initial preference to accept attractive candidates before the evidence accumulation process begins. As such, a plausible intervention would be delaying information about attractiveness until after participants have started viewing candidates' qualifications.

The Supplemental Materials outline three studies that investigated whether delaying face presentation impacted the relative start point. In Supplemental Study 1, participants completing a JBT where faces did not appear for 500 ms still showed a start point bias favoring more physically attractive applicants ($d = 0.19$), with similar effects emerging when the delay was increased to 1000 ms ($d = 0.19$) in Supplemental Study 2 or Supplemental Study 3 ($d = 0.11$). Each study also found criterion biases favoring more over less physically attractive applicants in conditions that delayed the presentation faces. In sum, 500-1000 ms delays were insufficient to reduce discrimination in observed choices or the relative start point.

The inability for delayed presentation to impact the start point bias may be due to the longer time course of JBT decisions; participants may not attend to the stimulus until the face appears.

Another possibility is that these initial biases result from the JBT's encoding phase that was used in Studies 1–2, where participants first passively view each application (including the candidate's face). However, all of the supplemental studies removed faces from the encoding phase and still showed start point biases in attractiveness. Though the present work is ambiguous as to why the relative start point bias occurs, these results indicate that delaying the onset of social information does not necessarily translate into lesser discrimination, nor does reducing discrimination require minimizing relative start point bias.

Finally, though the supplemental studies did not find that delaying the presentation of faces impacted behavioral outcomes or DDM parameters, these data may make for a productive application of "two-stage" decision models (Diederich & Busemeyer, 2006; Diederich & Trueblood, 2018). In these models, participants process one source of information (e.g., a face) before another source of information is given (e.g., relevant qualifications), and evidence can accumulate at different rates for each source. These two-stage models may arrive at different predictions than the analyses presented here if judgments are made primarily based on an applicant's physical appearance rather than their qualifications. Although outside the scope of the current work, our open data facilitate future investigations into this issue.

8. Conclusion

The present work reveals how discrimination can occur due to imperfect accuracy and the use of social information in judgment. Intervention strategies had unique impacts on the decision process, and DDM analyses clarified how these strategies reduced or exacerbated the impact of social information.

Subsequent research may look to other methods to validate and extend these findings. For instance, the drift rate results suggest participants completing the standard JBT use attractiveness information throughout the judgment process, but it is unclear whether this was due to attractiveness being treated as a qualification itself or the presence of attractiveness simply made the relevant qualifications appear more impressive. If attractiveness is treated as a qualification, participants may look at that information consistently while making their judgments. However, if an initial viewing of an attractive face only colors the interpretation of the outcome-relevant qualifications, this would be less likely to occur.

Eye-tracking analyses may then be helpful in teasing apart these competing accounts. Similarly, eye-tracking could further reveal how various interventions translate into behavior; for example, the awareness intervention used here may derive its effectiveness by reducing the duration participants focus on applicants' faces. In sum, these efforts expand our understanding of how discrimination emerges and the psychological processes that must be changed to reduce it.

Data availability

All data and materials are available on the Open Science Framework (https://tinyurl.com/jbtddm). All measures, manipulations, and exclusions are disclosed.

Declaration of Competing Interest

This research was partly supported by Project Implicit. J. R. Axt is Director of Data and Methodology for Project Implicit, Inc., a nonprofit organization with the mission to "develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race,

gender, or other factors."

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2021.104134.

References

Axt, J. R., Casola, G., & Nosek, B. A. (2019). Reducing social judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin, 45*(8), 1232–1251.

Axt, J. R., & Lai, C. K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology, 117*(1), 26–49.

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Brannon, S. M., Carr, S., Jin, E. S., Josephs, R. A., & Gawronski, B. (2019). Exogenous testosterone increases sensitivity to moral norms in moral dilemma judgements. *Nature Human Behavior, 3*, 856–866.

Carnes, M., Devine, P. G., Isaac, C., Manwell, L. B., Ford, C. E., Byars-Winston, A., … Sheridan, J. (2012). Promoting institutional change through bias literacy. *Journal of Diversity in Higher Education, 5*(2), 63.

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebelec, R. W., Massey, C., Duckworth, A. L., & Grant, A. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences, 116*(16), 7778–7783.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of Personality and Social Psychology, 108*(2), 219.

Devine, P. G., Forscher, P. S., Cox, W. T., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in stemm departments. *Journal of Experimental Social Psychology, 73*, 211–215.

Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics, 68*, 194–207.

Diederich, A., & Trueblood, J. S. (2018). A dynamic dual process model of risky decision making. *Psychological Review, 125*, 270–292.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*(4), 311–318.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception? *Psychological Review, 105*(3), 482.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology, 72*, 133–146.

Jaeger, B., Sleegers, W. W., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology, 75*, 102125.

JBT. (2019). *Axt & Lai*.

Johnson, D. J., Hopwood, C. J., Cesario, J., & Pleskac, T. J. (2017). Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Social Psychological and Personality Science, 8*(4), 413–423.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54.

King, E. B., & Ahmad, A. S. (2010). An experimental field study of interpersonal discrimination toward muslim job applicants. *Personnel Psychology, 63*(4), 881–906.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin, 36*(4), 512–523.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology, 100*(6), 1678–1712.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*, 469–485.

Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology, 16*, 113–154.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250–256.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*(4), 565–584.

Pleskac, T. J., Cesario, J., & Johnson, D. J. (2018). How race affects evidence accumulation during the decision to shoot. *Psychonomic Bulletin & Review, 25*, 1301–1330. https://doi.org/10.3758/s13423-017-1369-6.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.

Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science, 64*(11), 4988–4995.

R Core Team. (2020). R: A language and environment for statistical computing [computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007.

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of Human Resources, 44*(3), 710–735.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology, 64*(4), 583–639.

Sun, H., & Gao, L. (2019). Lending practices to same-sex borrowers. *Proceedings of the National Academy of Sciences, 116*(19), 9293–9302.

Todd, A. R., Johnson, D. J., Lassetter, B., Neel, R., Simpson, A., & Cesario, J. (2021). Category salience and racial bias in weapon identification: A diffusion modeling approach. *Journal of Personality and Social Psychology, 120*(3), 672–693.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437.

Turner, M. A., & Ross, S. L. (2003). *Discrimination in metropolitan housing markets phase ii: Asians and pacific islanders. (299318)*.

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*(6), 474–480.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*(1), 44–62. https://doi.org/10.1037/a0021765.

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods, 46*(1), 15–28. https://doi.org/10.3758/s13428-013-0369-3.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*(1), 117.

Jordan Axt is an assistant professor at McGill University. His research explores how people form and express intergroup bias in attitudes and behavior.

David Johnson is a postdoctoral researcher at the Lab for Applied Social Science Research at the University of Maryland. His research employs computational models and secondary data analyses to study the psychological processes that underlie decisions