Can we reduce facial biases?

Jaeger, Bastian; Todorov, Alexander; Evans, Anthony; van Beest, Ilja

Published in:
Journal of Experimental Social Psychology

DOI:
10.1016/j.jesp.2020.104004

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Jaeger, B., Todorov, A., Evans, A., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology*, *90*, [104004]. https://doi.org/10.1016/j.jesp.2020.104004

Journal of Experimental Social Psychology 90 (2020) 104004

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions[☆]

Bastian Jaeger[a,*], Alexander T. Todorov[b], Anthony M. Evans[a], Ilja van Beest[a]

[a] Department of Social Psychology, Tilburg University, the Netherlands
[b] Department of Psychology, Princeton University, United States of America

ARTICLE INFO

Keywords:
Face perception
Bias
Trustworthiness
Legal sentencing
Intervention

ABSTRACT

Trait impressions from faces influence many consequential decisions even in situations in which decisions should not be based on a person's appearance. Here, we test (a) whether people rely on trait impressions when making legal sentencing decisions and (b) whether two types of interventions—educating decision-makers and changing the accessibility of facial information—reduce the influence of facial stereotypes. We first introduced a novel legal decision-making paradigm. Results of a pretest ($n = 320$) showed that defendants with an untrustworthy (vs. trustworthy) facial appearance were found guilty more often. We then tested the effectiveness of different interventions in reducing the influence of facial stereotypes. Educating participants about the biasing effects of facial stereotypes reduced explicit beliefs that personality is reflected in facial features, but did not reduce the influence of facial stereotypes on verdicts (Study 1, $n = 979$). In Study 2 ($n = 975$), we presented information sequentially to disrupt the intuitive accessibility of trait impressions. Participants indicated an initial verdict based on case-relevant information and a final verdict based on all information (including facial photographs). The majority of initial sentences were not revised and therefore unbiased. However, most revised sentences were in line with facial stereotypes (e.g., a guilty verdict for an untrustworthy-looking defendant). On average, this actually *increased* facial bias in verdicts. Together, our findings highlight the persistent influence of trait impressions from faces on legal sentencing decisions.

People spontaneously infer a wide range of characteristics from a person's facial appearance. Demographic characteristics, such as a person's sex, age, or race, are perceived with near-perfect accuracy (Bruce & Young, 2012). Even perceptually ambiguous categories, such as sexual identity, social class, or political orientation can be detected at rates higher than chance (Alaei & Rule, 2016; Tskhay & Rule, 2013). People also infer personality traits from facial appearance (Todorov, Said, Engell, & Oosterhof, 2008). Stereotypes regarding what a trustworthy or competent person looks like are widely shared, but evidence for their accuracy is mixed at best (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Some studies suggest that personality impressions contain a small "kernel of truth" (Berry, 1990; Penton-Voak, Pound, Little, & Perrett, 2006). For example, a series of studies by Bonnefon, Hopfensitz, and De Neys (2013, 2017) showed that people can judge the trustworthiness of potential interaction partners at rates slightly higher than chance (ca. 55%). However, other studies found no accuracy in trustworthiness impressions (Efferson & Vogt, 2013; Rule, Krendl, Ivcevic, & Ambady, 2013) and evidence on the accuracy of other personality trait impressions (e.g., extraversion) is also mixed (Ames, Kammrath, Suppes, & Bolger, 2010; Jones, Kramer, & Ward, 2012; Kramer & Ward, 2010; Penton-Voak et al., 2006; Shevlin, Walker, Davies, Banyard, & Lewis, 2003). Although more research is needed to determine which personality traits can be judged with some level of accuracy, the current evidence suggests that people's ability to infer personality traits from faces is very limited at best.

Yet, research has shown that facial stereotypes guide many consequential decisions such as personnel selection, voting behavior, and economic exchange (Olivola, Funk, & Todorov, 2014). People even rely on trait impressions when more diagnostic cues are available (Olivola, Tingley, & Todorov, 2018; Rule, Bjornsdottir, Tskhay, & Ambady, 2016; Rule, Tskhay, Freeman, & Ambady, 2014), and when there are explicit rules that proscribe relying on a person's physical appearance (e.g., in legal sentencing; Wilson & Rule, 2015, 2016; Zebrowitz & McDonald, 1991). This overreliance on trait impressions from faces can lead to worse outcomes for decision-makers (Olivola & Todorov, 2010), but also to systematic discrimination against people with a certain facial

[☆] This paper has been recommended for acceptance by Nicholas Rule.

[*] Corresponding author at: Department of Social Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, the Netherlands.
E-mail address: bxjaeger@gmail.com (B. Jaeger).

https://doi.org/10.1016/j.jesp.2020.104004
Received 6 November 2019; Received in revised form 30 April 2020; Accepted 7 May 2020

appearance. For instance, competent-looking people are favored as business leaders, even though they do not seem to perform better (Graham, Harvey, & Puri, 2017; Ling, Luo, & She, 2019; Stoker, Garretsen, & Spreeuwers, 2016). Trustworthiness impressions predict capital punishment rulings despite their questionable accuracy (Wilson & Rule, 2015, 2016). In short, people appear to overrely on facial stereotypes when making a wide range of important decisions. As a consequence, researchers have called for efforts to mitigate the biasing influence of facial stereotypes (Olivola et al., 2014; Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015). Here, we answer this call by exploring the effectiveness of two types of interventions in reducing reliance on facial stereotypes.

1. Facial stereotypes influence decision-making

While there are numerous studies demonstrating the effects of facial stereotypes, comparatively little is known about *why* people persistently rely on trait impressions from faces. Addressing this question is crucial, as an understanding of the underlying mechanism not only advances theory, but is also a requirement for designing effective interventions. Recently, two (non-mutually exclusive) hypotheses have been put forward to address this gap. One explanation posits that the widespread influence of trait impressions can be explained by lay beliefs in the diagnostic value of facial appearance for inferring personality traits (Jaeger, Evans, Stel, & van Beest, 2019b; Rezlescu, Duchaine, Olivola, & Chater, 2012; Todorov, 2017). Many people believe in physiognomy—the idea that personality traits are reflected in an individual's facial appearance (Jaeger et al., 2019b; Suzuki, Tsukamoto, & Takahashi, 2017). Such beliefs may drive reliance on facial stereotypes because how much people rely on a certain cue is usually not determined by how predictive the cue actually is (i.e., how accurate trait impressions are), but by how predictive people *think* the cue is (i.e., how accurate people think their trait impressions are; Brunswik, 1956; Hammond, Hursch, & Todd, 1964). In fact, individual differences in physiognomic belief predict reliance on trait impressions when making economic trust decisions (Jaeger et al., 2019b): People who more strongly believe that trustworthiness is reflected in facial features rely more on their counterpart's perceived trustworthiness when deciding whom to trust. Thus, reliance on trait impressions may be driven by beliefs in the diagnostic value of facial appearance for judging an individual's personality.

A second explanation posits that the intuitive accessibility of trait impressions from faces can account for their persistent effects (Jaeger, Evans, Stel, & van Beest, 2019a). Faces attract attention (Ro, Russell, & Lavie, 2001; Theeuwes & Van der Stigchel, 2006) and are processed quickly and efficiently (Stewart et al., 2012; Willis & Todorov, 2006). This processing advantage leads to an intuitive accessibility of trait impressions from faces. As a consequence, reliance on facial stereotypes is relatively fast and not influenced by the restriction of cognitive capacities (Bonnefon et al., 2013; Jaeger et al., 2019a; Mieth, Bell, & Buchner, 2016). Crucially, previous research has shown that people favor readily available cues as they reduce decision effort (Evans & Krueger, 2016; Gigerenzer, Hertwig, & Pachur, 2011; Shah, 2007; Shah & Oppenheimer, 2008). Thus, people may rely on trait impressions from faces because it allows them to make decisions relatively effortlessly.

2. Reducing reliance on facial stereotypes

To sum up, previous research suggests that the pervasive influence of facial stereotypes is driven by a combination of (a) beliefs in the diagnostic value of facial appearance for inferring personality traits and (b) the intuitive accessibility of trait impressions from faces. Crucially, similar mechanisms have been identified in other research areas that investigate decision biases. Theories in the field of judgment and decision-making often distinguish between two general sources of bias: false beliefs (i.e., misconceptions) and automatically activated associations (i.e., misleading intuitions; Morewedge & Kahneman, 2017; Soll, Milkman, & Payne, 2014; Wilson & Brekke, 1994). Moreover, social psychological theories of bias typically distinguish between explicit and implicit expressions of bias (Devine, 1989; Dovidio, Kawakami, & Gaertner, 2002; Greenwald & Banaji, 1995; Greenwald, McGhee, Jordan, & Schwartz, 1998). Due to these similarities, we draw on the extensive literature on debiasing techniques in judgment and decision-making (Morewedge et al., 2015; Soll et al., 2014) and social psychology (Forscher et al., 2019; Lai et al., 2014) to design interventions aimed at reducing reliance on facial stereotypes.

A prominent strategy for reducing biases caused by misconceptions is to challenge beliefs through education (Chan, Jones, Hall Jamieson, & Albarracín, 2017; Soll et al., 2014). For example, educating people about compound interest can increase saving behavior (McKenzie & Liersch, 2011), educating people about cognitive biases can lead to more rational clinical decision-making (Hershberger, Markert, Part, Cohen, & Finger, 1997), and raising awareness of prejudice based on social group affiliation can reduce discrimination (Axt, Casola, & Nosek, 2018). Directly confronting participants with their stereotypes—rather than just raising awareness about the existence of stereotypes in general—has also been shown to reduce biased behavior (Czopp, Monteith, & Mark, 2006; Parker, Monteith, Moss-Racusin, & Van Camp, 2018). In Study 1, we therefore test whether we can reduce reliance on trait impressions by educating people about the influence of facial stereotypes or by confronting them with the fact that their facial stereotypes are not accurate.

A prominent strategy for reducing biases caused by intuitively available information is to design decision environments in such a way that participants are nudged to rely on the "right" cues (Soll et al., 2014; Thaler & Sunstein, 2008). The primary and efficient processing of faces leads to a quick availability of face-based impressions (Freeman & Johnson, 2016; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Crucially, information that is available first often exerts a disproportionate influence on decisions (Asch, 1946; Dimov & Link, 2017; Sullivan, 2018). Initial response tendencies are not sufficiently adjusted based on subsequently processed information (producing anchoring effects; Tamir & Mitchell, 2012; Tversky & Kahneman, 1974) and people are sometimes not able or willing to exert the cognitive effort required to integrate all available information (Shah & Oppenheimer, 2008; Simon, 1955). As a consequence, people often make decisions based on the cue that was processed first in order to reduce decision effort (Gigerenzer et al., 2008). This implies that manipulating how deeply and in which order information is processed could reduce the influence of facial stereotypes (Ghaffari & Fiedler, 2018). In Study 2, we therefore test whether preventing the primary processing of faces by presenting information sequentially (with faces being displayed after more relevant information) reduces reliance on facial stereotypes. We also test the effectiveness of prompting participants to make reflective rather than intuitive decisions.

We are not the first to test how different factors influence reliance on facial stereotypes. Providing information on how trustworthy a person has been in the past (Rezlescu et al., 2012) or giving feedback about a person's trustworthiness in a repeated interaction (Yu, Saleem, & Gonzalez, 2014) has been shown to reduce reliance on facial trustworthiness. In a similar vein, simply omitting photos from the decision-making environment would obviously eliminate any influence of facial appearance. These strategies may be effective, but they are not viable interventions in most real-world situations. When deciding on the culpability of a defendant or on the suitability of a job candidate, decision-makers are often faced with a limited amount of ambiguous or contradicting pieces of information, and it may not be possible to provide additional information about past behavior. It might also not be possible to completely remove information about a person's appearance. For these reasons, and in contrast to previous work, we focused on interventions that do not omit or add any additional decision-relevant information. Our goal was to test the effectiveness of different interventions under conditions that resemble the real-world situations in which the biasing effect of trait impressions from faces is particularly prevalent and problematic (e.g., in criminal sentencing, personnel selection, or voting).

3. The current studies

Here, we examine the effectiveness of different interventions in reducing the effect of facial stereotypes on legal sentencing decisions. We focus on decision-making in a legal context, because sentencing decisions can be immensely consequential, making biased decision-making particularly problematic. Appearance-based stereotyping undermines people's right to a fair trial (Lown, 1977). Yet, a host of studies has shown that facial stereotypes influence many real-life legal outcomes (Berry & Zebrowitz-McArthur, 1988; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Porter et al., 2010; Wilson & Rule, 2015, 2016; Zebrowitz & McDonald, 1991).

Similar to Zebrowitz and McDonald (1991), we focus on sentencing decisions in small claims court. Small claims court judges hear civil cases in which people can sue private citizens for relatively small amounts of money (e.g., up to $5000; the exact amount varies across countries). Plaintiffs and defendants often represent themselves and the evidence presented to the judge tends to be limited. However, the burden of proof is also relaxed in small claims cases: Plaintiffs do not need to present evidence that implicates the defendant “beyond reasonable doubt”, but judges rule in favor of the party that presents the most credible and convincing arguments. Given that small claims rulings reflect a more subjective interpretation of the evidence by the judge, it is possible that sentences are influenced by facial stereotypes. In fact, Zebrowitz and McDonald (1991) showed that babyfacedness—a facial feature that is correlated with perceived trustworthiness (Berry & Zebrowitz McArthur, 1986; Zebrowitz & Montepare, 1992)—predicted outcomes of small claims court rulings. Babyfaced defendants were found guilty less often (although this effect was only found for cases involving intentional, rather than negligent actions). Furthermore, when facing a babyfaced plaintiff, defendants that were found guilty had to pay a smaller fraction of the damages when they looked more babyfaced themselves. These results suggest that babyfaced individuals, who are generally seen as trustworthy, honest, and kind (Berry & Zebrowitz McArthur, 1986; Zebrowitz & Montepare, 1992), experience more leniency in court.

We present the results of three studies. All data, materials, preregistrations, and analysis scripts are available at the Open Science Framework (https://osf.io/h4yf3/). We report how our sample sizes were determined, all data exclusions, and all measures. In a pretest (n = 320), we develop and validate a legal sentencing paradigm that measures reliance on facial stereotypes. We examine whether the facial trustworthiness of plaintiffs and defendants influences sentencing decisions in small claims court cases. We then test the effectiveness of two types of interventions in reducing reliance on trait impressions in two preregistered studies. In Study 1 (n = 979), we educate participants about the low diagnostic value of facial appearance for inferring personality traits. In Study 2 (n = 975), we change the decision-making environment to disrupt the intuitive accessibility of trait impressions.

4. Pretest

We created a novel legal sentencing task, tailored to measure reliance on facial stereotypes. Previous experimental studies have predominantly taken two methodological approaches. In some studies, participants view a series of face images and indicate perceptions of culpability or sentencing decisions (e.g., Wilson & Rule, 2016). Multiple trials with within-subjects manipulations of facial appearance increase statistical power, but providing little or no background information on the cases limits the ecological validity of the task. In other studies, participants receive realistic case descriptions including relevant extenuating or aggravating facts (e.g., Berry & Zebrowitz-McArthur, 1988; Gunnell & Ceci, 2010). This approach more closely resembles the conditions in which decisions are made in real life. However, these studies usually consist of between-subject designs with few cases and face images, limiting statistical power and the generalizability of the results.

Here, we tried to incorporate advantages of the two approaches. Based on descriptions of real small claims court cases, we created ten fictitious case files, with plaintiffs filing suits against defendants. Cases included realistic evidence and we manipulated the perceived trustworthiness of plaintiffs and defendants in a within-subjects design. Participants indicated sentencing decisions for all ten cases. In line with previous studies (Berry & Zebrowitz-McArthur, 1988; Wilson & Rule, 2016), we expected participants to find defendants guilty more often when they look untrustworthy (vs. trustworthy). We also measured confidence in verdicts and, in case participants ruled in favor of the plaintiff, the damages they wished to award to the plaintiff. This allowed us to explore whether congruence between sentences and facial stereotypes (e.g., a guilty verdict for untrustworthy defendants) would increase confidence in verdicts. Moreover, we explored whether untrustworthy-looking defendants are punished twice, by being more likely to be found guilty and by receiving a harsher sentence (i.e., being ordered to pay more damages).

4.1. Methods

4.1.1. Participants

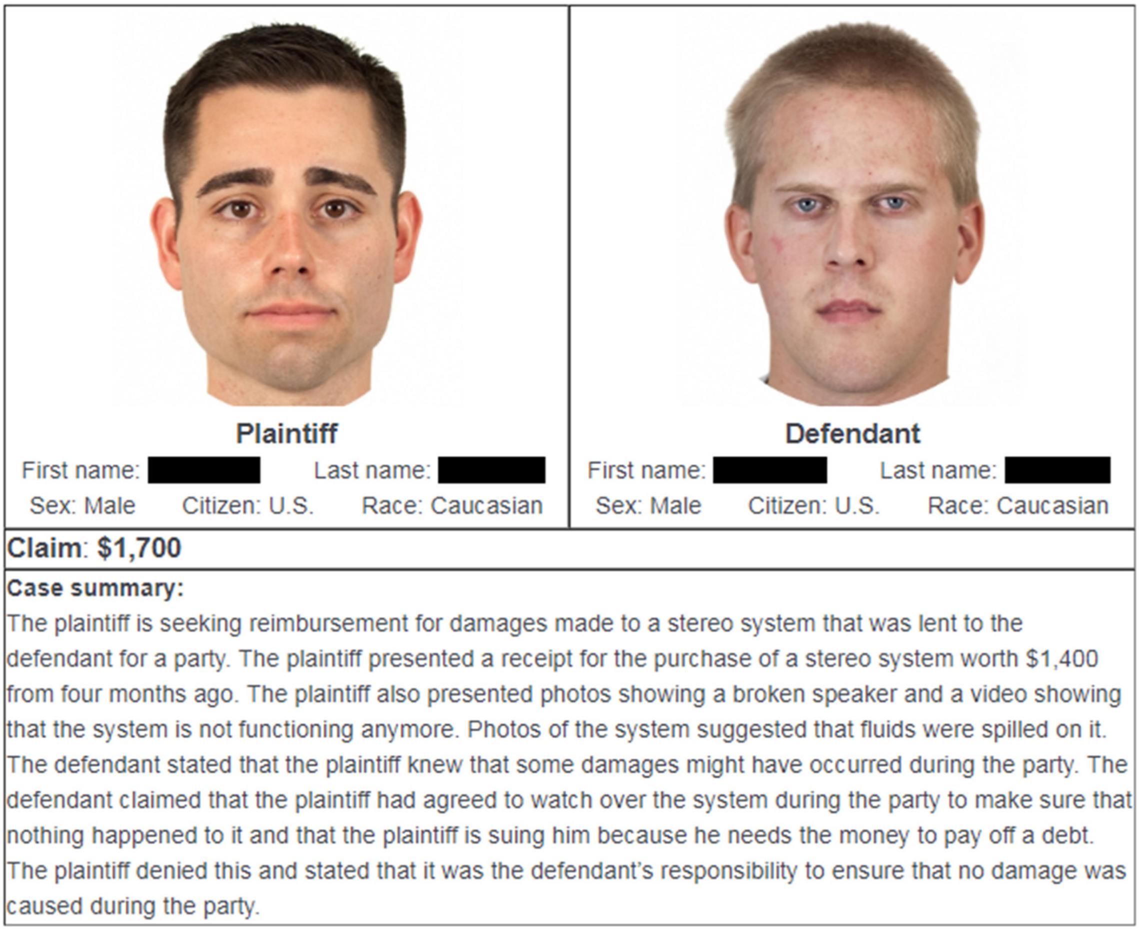
We recruited a total of 363 U.S. American workers from Amazon Mechanical Turk (MTurk; Paolacci & Chandler, 2014) who participated in exchange for $1.50. Data from 30 participants (8.26%) who failed an attention check at the end of the study and 8 participants (2.40%) who indicated having only a poor or basic English proficiency were excluded from analysis, leaving a final sample of 325 participants (50.46% female, M_{age} = 35.91, SD_{age} = 10.03).

4.1.2. Materials

We created case files for ten fictitious small claims court cases (see Fig. 1). Case files included a photo and demographic information on the plaintiff and the defendant. All individuals were White male U.S. citizens and had their first and last name redacted. Case files also included the size of the plaintiff's claim (ranging from $600 to $3600) and a case summary of approximately 130 words. Each summary mentioned the reason why the plaintiff was suing the defendant (e.g., seeking reimbursement for a damaged stereo system) and the evidence that was presented by the plaintiff and the defendant (e.g., photos of a broken speaker, a receipt confirming the purchase of a stereo system). In line with real-world small claims court cases, the evidence presented by both sides was relatively limited.

We selected 20 images of White male individuals from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The database includes ratings of all targets on various trait dimensions. We selected the ten individuals who received the lowest (M = 2.62, SD = 0.17) and highest (M = 3.78, SD = 0.09) ratings on perceived trustworthiness. Targets varied in perceived age with average age ratings ranging from 19.5 to 43.2 years (M = 28.60, SD = 6.90). Age ratings of the trustworthy-looking targets (M = 28.67, SD = 6.64) and untrustworthy-looking targets (M = 28.57, SD = 7.50) were very similar.

Next, we manipulated the perceived trustworthiness of all targets. Oosterhof and Todorov (2008) created a series of computer-generated face prototypes that reflect the typical facial appearance of targets varying on several trait dimensions (e.g., trustworthiness, dominance). We selected two face prototypes that reflect a high (i.e., three standard deviations above the mean) and low (i.e. three standard deviations below the mean) score on perceived trustworthiness. Using Psychomorph (Tiddeman, Burt, & Perrett, 2001), we transformed each target's face shape towards the face shape of the computer-generated prototype by 60%. Trustworthy-looking targets were morphed with the



Plaintiff

First name: [redacted] Last name: [redacted]

Sex: Male Citizen: U.S. Race: Caucasian

Defendant

First name: [redacted] Last name: [redacted]

Sex: Male Citizen: U.S. Race: Caucasian

Claim: **$1,700**

Case summary:

The plaintiff is seeking reimbursement for damages made to a stereo system that was lent to the defendant for a party. The plaintiff presented a receipt for the purchase of a stereo system worth $1,400 from four months ago. The plaintiff also presented photos showing a broken speaker and a video showing that the system is not functioning anymore. Photos of the system suggested that fluids were spilled on it. The defendant stated that the plaintiff knew that some damages might have occurred during the party. The defendant claimed that the plaintiff had agreed to watch over the system during the party to make sure that nothing happened to it and that the plaintiff is suing him because he needs the money to pay off a debt. The plaintiff denied this and stated that it was the defendant's responsibility to ensure that no damage was caused during the party.

Fig. 1. A case file with a trustworthy-looking plaintiff and an untrustworthy-looking defendant.

trustworthy-looking face prototype, whereas untrustworthy-looking targets were morphed with the untrustworthy-looking face prototype. This procedure somewhat exaggerated the facial features linked to perceptions of trustworthiness and allowed us to create prototypically (un-)trustworthy-looking individuals without compromising the realistic nature of the face stimuli.

Finally, we matched case files and face images. Each case featured a plaintiff and a defendant differing on perceived trustworthiness: One individual looked trustworthy while the other looked untrustworthy. We created four sets of stimuli. Each set contained all ten case files and all 20 face images. In each set, face images were randomly matched to a case and a role (i.e., plaintiff or defendant). Half of all cases featured a trustworthy-looking plaintiff and an untrustworthy-looking defendant, while the roles were reversed in the other half.

4.1.3. Procedure

Participants were randomly assigned to one of the four stimulus sets. To measure sentencing decisions, participants were instructed to carefully read each case and to indicate a sentence by ruling in favor of the plaintiff or the defendant. After each ruling, participants also indicated their confidence in the ruling on a scale that ranged from 1 (*not confident at all*) to 9 (*extremely confident*). In case participants ruled in favor of the plaintiff, they were asked to indicate the amount of damages that the plaintiff should be awarded on a scale that ranged from 50% to 100% (in steps of 10%) of the original claim.

4.1.4. Sensitivity analysis

We conducted a post hoc sensitivity analysis to determine the smallest effect size we were able to detect for our main effect of interest (the effect of facial trustworthiness on verdicts) with 80% power (and α = 5%). As software commonly used for sensitivity analyses, such as G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), does not support multilevel data, we relied on the *simr* package (Green & Macleod, 2016) in R (R Core Team, 2019). The package provides power estimates for fixed effects in multilevel regression models. We systematically varied the effect of facial trustworthiness on verdicts and calculated power at each level, to test which effect size we were able to detect with at least 80% power. This showed that we had 80% power to detect an odds ratio of 1.27 for the effect of facial trustworthiness on verdict. To illustrate, an odds ratio of this size corresponds to a six percentage point difference in guilty verdicts (e.g., 50% vs. 56%) for trustworthy-looking versus untrustworthy-looking defendants.

4.2. Results

On average, participants found the defendant guilty 53.26% of the time (*SD* = 18.54%). Two participants (0.62%) found all defendants guilty, whereas three (0.92%) found none guilty. The prevalence of guilty verdicts varied across cases (*Min* = 34.15%, *Max* = 71.08%, *M* = 53.26%, *SD* = 13.12%).

We analyzed the effect of facial trustworthiness on sentencing decisions by estimating a multilevel regression model with random intercepts and slopes per participant and per case. This accounts for variation in the overall rate of guilty verdicts across participants (i.e., some participants indicating more guilty verdicts than others) and across cases (i.e., defendants in some cases receiving more guilty verdicts than others). Regressing verdict (0 = defendant is not guilty, 1 = defendant is guilty) on facial trustworthiness

(-0.5 = trustworthy-looking defendant, 0.5 = untrustworthy-looking defendant) revealed a positive effect, $\beta = 0.319$, $SE = 0.080$, $z = 3.94$, $p < .001$, 95% CI [0.161, 0.477], $OR = 1.38$. The rate of guilty verdicts was 8.03 percentage points higher for untrustworthy-looking defendants (56.65% vs. 48.61%).

We also explored whether defendants' facial trustworthiness affected confidence in verdicts or the amount of money participants awarded to the plaintiff in case of a guilty verdict. Regressing confidence on facial trustworthiness, verdict, and their interaction showed a positive effect of a guilty verdict, $\beta = 0.243$, $SE = 0.052$, $t(3016) = 4.75$, $p < .001$, 95% CI [0.142, 0.344]. Participants were more confident in their verdicts when they ruled in favor of the plaintiff. There was no effect of facial trustworthiness, $\beta = -0.020$, $SE = 0.071$, $t(1{,}099) = 0.29$, $p = .77$, 95% CI [-0.159, 0.118], and no interaction effect between verdict and facial trustworthiness, $\beta = 0.088$, $SE = 0.099$, $t(2990) = 0.89$, $p = .37$, 95% CI [-0.106, 0.281].

Finally, regressing the amount of money that was awarded to the plaintiff in case of a guilty verdict on facial trustworthiness revealed a small positive effect, $\beta = 1.655$, $SE = 0.727$, $t(137.4) = 2.28$, $p = .024$, 95% CI [0.191, 3.078]. Participants awarded the plaintiff 1.63 percentage points more of their original claim when the defendant looked untrustworthy (85.28% vs. 83.64%).

4.3. Discussion

Results of the pretest showed that legal sentencing decisions were influenced by the facial trustworthiness of the involved parties. The rate of guilty verdicts was 8.03 percentage points higher when the defendant looked untrustworthy (vs. trustworthy). Facial trustworthiness also influenced how much money participants awarded to the plaintiff in case of a guilty verdict, with plaintiffs receiving 1.63 percentage points more when they were suing an untrustworthy-looking (vs. trustworthy-looking) defendant. We did not find any evidence that confidence in verdicts was influenced by facial trustworthiness. Thus, using a novel sentencing task with multiple cases and controlled manipulations of facial trustworthiness, we replicate prior work showing that people rely on trait impressions from faces when making legal sentencing decisions (Porter et al., 2010; Wilson & Rule, 2015, 2016). Our findings also replicate previous work by Zebrowitz and McDonald (1991) who found that babyfacedness—a facial feature that is correlated with perceived trustworthiness (Berry & Zebrowitz McArthur, 1986; Zebrowitz & Montepare, 1992) —influenced verdicts and awarded damages in real-world small claims cases.

5. Study 1: belief interventions

In Study 1, we used the sentencing task that was developed and validated in the pretest to test the effectiveness of an intervention in reducing reliance on facial trustworthiness. Our goal was to reduce reliance on facial stereotypes by reducing explicit beliefs that personality can be judged from facial appearance (Jaeger et al., 2019b). In one condition, participants read a text that informed them about scientific research on facial stereotypes. The text mentioned the automatic accessibility of facial stereotypes, that facial stereotypes are usually not accurate, and that relying on them can result in worse decision-making outcomes. The intervention specifically focused on facial stereotypes, as previous work suggests that raising awareness of stereotypes in general may not be effective (Axt et al., 2018). Our manipulation was modelled after previous research in the domain of lay beliefs. For instance, Levy, Stroessner, and Dweck (1998) used fake scientific articles to manipulate beliefs in the innateness of personality traits and this influenced how strongly participants associated different social groups with stereotypical personality traits.

In a second intervention condition, we additionally confronted participants with the low diagnostic value of their facial stereotypes. Before reading the educational text, we showed participants ten pairs of faces. Their task was to identify which of the two individuals was a convicted felon. We told participants that they only guessed four out of ten correctly, meaning that their guesses were not better than chance. We measured physiognomic beliefs (i.e., participants' explicit beliefs that personality traits can be judged accurately from faces) in all conditions and hypothesized that, compared to a control condition in which participants were not exposed to a manipulation, both interventions would reduce physiognomic beliefs and reliance on facial trustworthiness when making sentencing decisions.

5.1. Methods

5.1.1. Power analysis

We conducted an a priori power analysis using the *simr* package in R, which allows one to test how power varies as a function of the number of levels of a random effect (in our case, the number of participants or the number of cases). As the number of cases was fixed, we tested how power varies across different numbers of participants. Calculating power across a wide range of sample sizes showed that 250 participants per condition are required to detect a 30% decrease in the effect of facial trustworthiness on verdicts with 80% power (and $\alpha = 5\%$). As a conservative measure, we decided to recruit 325 participants per condition.

5.1.2. Participants

We recruited a total of 1249 US American workers from Amazon Mechanical Turk who participated in exchange for $2.50. Data from 227 participants (18.17%) who failed an attention check at the end of the study and from 42 participants (4.11%) who indicated poor or basic English proficiency were excluded from analysis, leaving a final sample of 979 participants (47.40% female, $M_{\text{age}} = 36.14$, $SD_{\text{age}} = 11.24$).

5.1.3. Materials & procedure

Participants were randomly allocated to one of three conditions. In all conditions, participants completed the legal sentencing task as described in the previous study. For each case, they ruled in favor of the plaintiff or the defendant and indicated their confidence in the ruling on a scale that ranged from 1 (*not confident at all*) to 9 (*extremely confident*). Next, to measure belief in the visibility of personality traits in facial appearance, participants completed the physiognomic belief scale (Jaeger et al., 2019b). Participants were prompted to imagine seeing the passport photo of a stranger. They were asked to indicate how much they agree with three statements (e.g., *I can learn something about a person's personality just from looking at his or her face*) on a scale from 1 (strongly disagree) to 7 (strongly agree). Average scores across the three items constituted our measure of physiognomic beliefs (Cronbach's $\alpha = 0.84$).

The three conditions only differed in the texts participants were exposed to prior to completing the sentencing task. In the education condition ($n = 332$), participants read an educational text about personality impressions from faces that was approximately 300 words long. First, participants were told that people spontaneously form impressions of others' personality based on their facial appearance; that there is substantial agreement on what, for example, a trustworthy person looks like; and that these judgments are formed very quickly, sometimes without the perceiver's awareness. To illustrate these points, participants were shown two face images of a typical trustworthy-looking and untrustworthy-looking face (drawn from a database of computer-generated faces varying in perceived trustworthiness; Oosterhof & Todorov, 2008). Next, the text mentioned that trustworthiness impressions influence many important decisions even though research suggests that these impressions are often inaccurate. It was also highlighted that this is problematic because it leads to unfair treatment of people with a certain facial appearance (the exact text can be found in the online materials).

In the education-and-confrontation condition (n = 332), prior to reading the educational text, participants completed an additional task that was designed to demonstrate that their face-based impressions are inaccurate. Participants saw ten pairs of faces of male individuals that were taken from the 10k Faces Database (Bainbridge, Isola, Blank, & Oliva, 2013). Participants were told that each pair included one convicted felon and that their task was to identify that person. Feedback about accuracy was standardized across all participants. They were told that they only guessed four out of ten correctly, meaning that their guesses were not better than chance.

In the control condition (n = 315), participants read a text about the geography of Scotland.

After reading the respective texts, participants answered three comprehension check questions (e.g., *research shows that first impressions influence many important decisions*). Participants could only proceed to the sentencing task after having answered all three questions correctly.

5.2. Results

Participants found the defendant guilty 51.47% of the time (SD = 17.47%). Three participants (0.31%) found all defendants guilty, whereas four (0.41%) found none guilty. The prevalence of guilty verdicts varied across cases (Min = 36.26%, Max = 63.53%, M = 51.47%, SD = 10.40%).

5.2.1. Physiognomic beliefs

First, we tested whether the interventions reduced beliefs that personality is reflected in facial appearance. Compared to participants in the control condition (M = 3.80, SD = 1.37), participants in the education condition (M = 3.59, SD = 1.33) indicated lower physiognomic beliefs, t(640.6) = 2.03, p = .042, d = 0.16, and so did participants in the education-and-confrontation condition (M = 3.50, SD = 1.38), t(643.5) = 2.80, p = .005, d = 0.22. Physiognomic beliefs did not significantly differ between the education and education-and-confrontation condition, t(661.2) = 0.82, p = .41, d = 0.06. These results show that both interventions were successful in reducing the belief that personality is reflected in facial features, although differences were relatively small.

5.2.2. Sentencing decisions

Next, we tested whether the interventions reduced reliance on facial trustworthiness in the legal sentencing task. We estimated a multilevel regression model with random intercepts and slopes per participant and case. Regressing verdict (0 = defendant is not guilty, 1 = defendant is guilty) on facial trustworthiness (−0.5 = trustworthy-looking defendant, 0.5 = untrustworthy-looking defendant), condition (with the control condition being the reference category), and their interaction terms revealed a positive effect of facial trustworthiness, β = 0.327, SE = 0.101, z = 3.22, p = .001, 95% CI [0.127, 0.527], OR = 1.39. There were no significant differences in guilty verdicts between the control condition and the education condition, β = 0.107, SE = 0.061, z = 1.74, p = .083 95% CI [−0.014, 0.227], OR = 1.11, or the education-and-confrontation condition, β = 0.059, SE = 0.061, z = 0.96, p = .34, 95% CI [−0.062, 0.179], OR = 1.06. The difference between the education condition and the education-and-confrontation condition was also not significant, β = −0.048, SE = 0.061, z = 0.79, p = .43, 95% CI [−0.167, 0.071], OR = 0.95.

Crucially, examining the interaction effects showed that, compared to the control condition, the effect of facial trustworthiness on verdicts was not significantly different in the education condition, β = 0.132, SE = 0.133, z = 0.99, p = .32, 95% CI [0.130, 0.395], OR = 1.14, or in the education-and-confrontation condition, β = −0.056, SE = 0.134, z = 0.42, p = .68, 95% CI [−0.318, 0.207], OR = 0.95 (see Fig. 2). The difference between the education condition and the education-and-confrontation condition was also not significant, β = −0.188, SE = 0.132, z = 1.42, p = .16, 95% CI [−0.448, 0.072], OR = 0.83.[1] Thus, neither intervention was successful in reducing reliance on facial trustworthiness. The rate of guilty verdicts was 7.78 percentage points higher for untrustworthy-looking defendants in the control condition (54.76% vs. 46.98%), 11.34 percentage points higher in the education condition (58.71% vs. 47.37%), and 6.65 percentage points higher in the education-and-confrontation condition (54.69% vs. 48.04%).

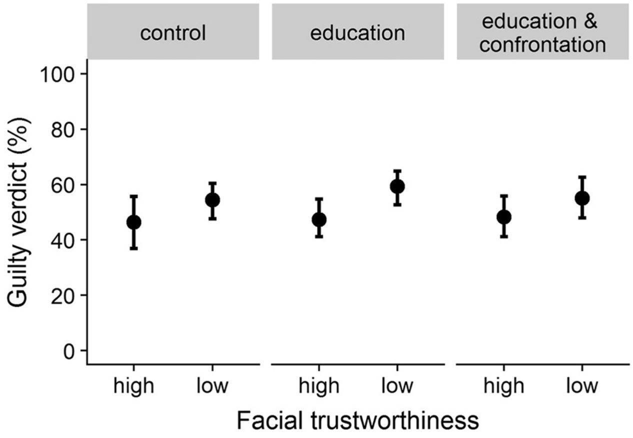


Fig. 2. Differences in rates of guilty verdicts for trustworthy-looking and untrustworthy-looking defendants as a function of condition. Dots denote predicted values. Error bars denote bootstrapped 95% confidence intervals.

5.2.3. Confidence in verdicts

We also tested whether the interventions influenced confidence in verdicts. Regressing confidence on facial trustworthiness, verdict, and condition yielded a positive effect of a guilty verdict, β = 0.180, SE = 0.030, t(9,076) = 5.99, p < .001, 95% CI [0.120, 0.238]. As in our pretest, participants were more confident in their verdicts when they found the defendant guilty. There was no effect of facial trustworthiness, β = 0.033, SE = 0.049, t(6.25) = 0.68, p = .52, 95% CI [−0.070, 0.136], and compared to the control condition, confidence was not significantly different in the education condition, β = −0.001, SE = 0.091, t(973.0) = 0.01, p = .99, 95% CI [−0.180, 0.178], or in the education-and-confrontation condition, β = −0.145, SE = 0.091, t (973.4) = 1.59, p = .11, 95% CI [−0.323, 0.034]. In other words, we did not find evidence that the interventions reduced confidence in verdicts.

5.2.4. Exploratory analyses

To further probe the effects of the two interventions, we conducted Bayesian analyses using the *BayesFactor* package (Morey & Rouder, 2018) in R (R Core Team, 2019). Bayesian t-tests with default Cauchy priors yielded substantial support for the null hypothesis of no difference between the control condition and the education condition, BF_{01} = 6.49, and strong support for the null hypothesis of no difference between the control condition and the education-and-confrontation condition, BF_{01} = 10.66. These results support the conclusion that neither intervention significantly reduced reliance on facial trustworthiness.

The interventions were based on a proposed link between belief in the visibility of personality in a person's facial appearance and reliance on trait impressions when making decisions (Jaeger et al., 2019b). Even though the interventions somewhat reduced physiognomic beliefs, they did not reduce reliance on facial trustworthiness, raising the question

[1] Comparing the control condition against a combination of both intervention conditions also yielded no significant difference in the effect of facial trustworthiness on verdicts, β = 0.038, SE = 0.116, z = 0.33, p = .74, 95% CI [−0.190, 0.267], OR = 1.04.

whether physiognomic beliefs were actually related to reliance on facial trustworthiness. To test this, we extracted participant-specific slopes for the effect of facial trustworthiness from our multilevel regression models, as an indicator of how much each participant relied on trait impressions when making sentencing decisions. There was indeed a significant correlation between physiognomic beliefs and reliance on facial trustworthiness, $r(977) = 0.200$, $p < .001$. There was also a positive correlation between physiognomic beliefs and confidence in verdicts, $r(977) = 0.204$, $p < .001$. Participants who more strongly endorsed the belief that personality is reflected in facial features relied more on facial trustworthiness when making sentencing decisions and they were more confident in their verdicts. These results rule out the explanation that the observed reduction in physiognomic beliefs did not translate to less biased sentencing decisions because there was no link between beliefs and behavior.

5.3. Discussion

Neither intervention successfully reduced the effect of facial stereotypes on sentencing decisions. Educating participants about the low accuracy of their trait impressions reduced explicit beliefs in the diagnostic value of facial appearance for inferring personality traits, but this effect was relatively small. Importantly, the intervention did not reduce reliance on facial stereotypes when making sentencing decisions and it did not reduce confidence in verdicts. The same pattern was observed for a second intervention: Even when participants were directly confronted with the low accuracy their trait impressions, they continued to rely on them when making decisions in a subsequent task.

6. Study 2: accessibility interventions

In Study 2, we tested the effectiveness of an alternative intervention in reducing reliance on facial trustworthiness. Trait impressions from faces are intuitively accessible (Stewart et al., 2012; Todorov et al., 2009; Willis & Todorov, 2006) and accessible information often exerts a disproportionate influence on decisions (Shah, 2007; Simmons & Nelson, 2006; Tversky & Kahneman, 1974). To disrupt the primary processing of faces, we presented information sequentially. First, participants saw only case-relevant information and indicated an initial verdict. Then, they saw the entire case file (which also included face images of the plaintiff and defendant) and indicated their final verdict. We hypothesized that the majority of participants would not revise their initial verdicts. Reliance on intuitively available trait impressions constitutes a low-effort decision strategy and people might not be aware of the extent to which their decisions are influenced by facial stereotypes (Jaeger et al., 2019a). In our sequential design, participants have to actively revise their verdict (and ignore case-relevant information) if they want to base their decisions on the parties' facial appearance. They might be reluctant to do so because sticking with their initial verdict should reduce decision effort (Shah & Oppenheimer, 2008; Simon, 1955). Any initial verdict that is not revised reflects a verdict that is unbiased by facial stereotypes, as participants were not exposed to face images when deciding on their initial verdicts. Thus, if the majority of initial verdicts are not overturned, this should reduce the overall influence of facial stereotypes on verdicts compared to a control condition in which participants do not make decisions sequentially and are exposed to the face images right away.

In a second intervention condition, we tested whether the influence of intuitively available trait impressions would be further reduced by prompting participants to make reflective decisions (Newman, Gibb, & Thompson, 2017). To ensure that initial verdicts are based on a careful consideration of the case-relevant information, participants had to reflect on their initial verdicts for at least 30 s before they could indicate their decision.

6.1. Methods

6.1.1. Participants

Based on the results of the power analysis reported in Study 1, we again decided to recruit 325 participants per condition. We recruited a total of 1085 U.S. American workers from Amazon Mechanical Turk who participated in exchange for $2.50. Data from 93 participants (8.57%) who failed an attention check at the end of the study and from 17 participants (1.71%) who indicated a poor or basic English proficiency were excluded from analysis, leaving a final sample of 975 participants (49.74% female, $M_{\text{age}} = 35.86$, $SD_{\text{age}} = 10.50$).

6.1.2. Materials & procedure

Participants were randomly allocated to one of three conditions. In all conditions, participants completed the legal sentencing task as described in our Pretest. For each case, they ruled in favor of the plaintiff or the defendant and indicated their confidence in the ruling on a scale that ranged from 1 (*not confident at all*) to 9 (*extremely confident*).

In the sequential condition ($n = 319$), participants first saw the case files without any personal information about the plaintiff or defendant and were asked to indicate an initial ruling in favor of the plaintiff or the defendant. Next, participants saw the entire case files, including the images of the plaintiff and defendant, and were asked to indicate their final ruling and their confidence in the ruling on a scale that ranged from 1 (*not confident at all*) to 9 (*extremely confident*).

In the sequential-and-reflection condition ($n = 329$), participants followed the same procedure as in the sequential condition, but they were prompted to think carefully and make reflective decisions for all cases (Newman et al., 2017). They could only indicate an initial ruling after 30 s had passed and they were instructed to take at least this long to carefully study the case summary before indicating a ruling.

In the control condition ($n = 327$), participants completed the legal sentencing task without the order of stimuli being manipulated.

6.2. Results

Participants found the defendant guilty 52.01% of the time ($SD = 16.44\%$). Five participants (0.51%) found all defendants guilty, whereas four (0.41%) found none guilty. The prevalence of guilty verdicts varied across cases ($Min = 31.01\%$, $Max = 69.23\%$, $M = 52.02\%$, $SD = 13.52\%$).

6.2.1. Response times

First, we analyzed response times for initial rulings to check whether instructions to reflect on decisions in the sequential-and-reflection condition actually led to longer decision times compared to the sequential condition. Response times were $\log_{10}$-transformed due to their right-skewed distribution. A t-test showed that participants in the sequential-and-reflection condition ($M = 1.658$, $SD = 0.111$) took longer to reach a decision compared to participants the sequential condition ($M = 1.527$, $SD = 0.273$), $t(417.5) = 7.93$, $p < .001$, $d = 0.62$.[2]

6.2.2. Sentencing decisions

Next, we tested whether our interventions reduced reliance on facial trustworthiness in the legal sentencing task. We estimated a multilevel regression model with random intercepts and slopes per participant and case. Regressing verdict (0 = defendant is not guilty, 1 = defendant is guilty) on facial trustworthiness (-0.5 = trustworthy-looking defendant, 0.5 = untrustworthy-looking defendant), condition (with the control condition being the reference category), and their interaction terms revealed a positive effect of facial trustworthiness, $\beta = 0.218$,

[2] Excluding 63 raw response times (0.65%) that were more than three standard deviations above the mean response time before $\log_{10}$-transforming the data produced a similar result.

SE = 0.105, *z* = 2.08, *p* = .038, 95% CI [0.005, 0.431], *OR* = 1.24. There were no significant differences in rates of guilty verdicts between the control condition (50.89%) and the sequential condition (52.97%), β = 0.095, *SE* = 0.058, *z* = 1.65, *p* = .10 95% CI [−0.018, 0.209], *OR* = 1.10, or the sequential-and-reflection condition (52.20%), β = 0.060, *SE* = 0.057, *z* = 1.05, *p* = .30, 95% CI [−0.053, 0.173], *OR* = 1.06. The difference between the sequential condition and the sequential-and-reflection condition was also not significant, β = −0.081, *SE* = 0.117, *z* = 0.69, *p* = .49, 95% CI [−0.310, 0.148], *OR* = 0.92.[3]

Crucially, we found that, compared to the control condition, the effect of facial trustworthiness on verdicts was significantly *larger* in the sequential condition, β = 0.529, *SE* = 0.116, *z* = 4.54, *p* < .001, 95% CI [0.301, 0.758], *OR* = 1.70, and in the sequential-and-reflection condition, β = 0.448, *SE* = 0.115, *z* = 3.88, *p* < .001, 95% CI [0.222, 0.675], *OR* = 1.56 (see Fig. 3). The difference between the sequential condition and the sequential-and-reflection condition was not significant, β = −0.081, *SE* = 0.117, *z* = 0.69, *p* = .49, 95% CI [−0.310, 0.148], *OR* = 0.92. Thus, contrary to our predictions, both interventions significantly *increased* the influence of facial trustworthiness. The rate of guilty verdicts was 5.40 percentage points higher for untrustworthy-looking defendants in the control condition (53.51% vs. 48.11% guilty verdicts), 18.40 percentage points higher in the sequential condition (62.25% vs. 43.85% guilty verdicts), and 16.81 percentage points higher the sequential-and-reflection condition (60.54% vs. 43.73% guilty verdicts).

6.2.3. *Confidence in verdicts*

We also tested whether the interventions influenced confidence in verdicts. Regressing confidence on facial trustworthiness, verdict, and condition yielded no effect of facial trustworthiness, β = 0.042, *SE* = 0.065, *t*(8.58) = 0.65, *p* = .54, 95% CI[−0.092, 0.177], but a positive effect of a guilty verdict, β = 0.086, *SE* = 0.030, *t* (9,023) = 2.81, *p* = .005, 95% CI [0.026, 0.145]. Participants were more confident in their verdicts when they found the defendant guilty. Compared to the control condition, confidence was significantly higher in the sequential condition, β = 0.333, *SE* = 0.093, *t*(971.8) = 3.59, *p* < .001, 95% CI [0.151, 0.515], and also in the sequential-and-reflection condition, β = 0.402, *SE* = 0.092, *t*(972.3) = 4.73, *p* < .001, 95% CI [0.222, 0.583]. There was no significant difference in confidence between the sequential condition and the sequential-and-reflection condition β = 0.070, *SE* = 0.093, *t*(972.8) = 0.75, *p* = .45, 95% CI [−0.112, 0.251]. Thus, the interventions significantly increased confidence in verdicts.

6.2.4. *Exploratory analyses*

To further probe the effects of the two interventions, we again conducted Bayesian analyses. Bayesian *t*-tests with default Cauchy priors yielded strong support for the alternative hypothesis that, compared to the control condition, reliance on facial trustworthiness was stronger in the sequential condition, BF_{10} = 1484, and in the sequential-and-reflection condition, BF_{10} = 188. These results support the conclusion that both interventions significantly increased reliance on facial trustworthiness.

Finally, we analyzed how often and under what conditions participants revised their initial decision to understand why the interventions increased rather than decreased reliance on facial trustworthiness. We hypothesized that most participants would not revise their initial decisions, which were made in the absence of face images and therefore unbiased by facial stereotypes. In fact, the majority of initial rulings in the sequential condition (89.78%) and in sequential-and-reflection condition (90.61%) were not revised when participants saw the images of the plaintiff and defendant and had the chance to change their verdict. However, analyzing revision rates showed that participants were more likely to revise their initial ruling when it was not in line with face stereotypes (e.g., a trustworthy-looking defendant being found guilty; 15.4%) than when it was already in line with stereotypes (3.14%), $\chi^2(1)$ = 310.2, *p* < .001. Of all revised rulings, 83.52% ended up being congruent with face stereotypes whereas only 16.48% were incongruent with face stereotypes. As a consequence, while only 51.11% of all initial rulings made in the absence of face images were in line with face stereotypes, 57.61% of all final rulings made in the presence of face images were, $\chi^2(1)$ = 55.12, *p* < .001. In sum, in the absence of face images, both interventions were successful in producing unbiased rulings, which were seldom revised when participants did have access to the face images. However, the wide majority of revisions that did occur brought decisions in line with face stereotypes. This increased the overall effect of facial trustworthiness on sentencing decisions.

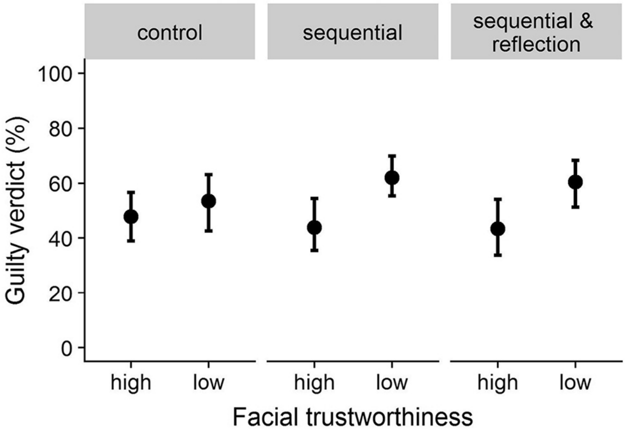


Fig. 3. Differences in rates of guilty verdicts for trustworthy-looking and untrustworthy-looking defendants as a function of condition. Dots denote predicted values. Error bars denote bootstrapped 95% confidence intervals.

6.3. *Discussion*

Results of Study 2 showed that both interventions increased, rather than decreased, reliance on facial stereotypes. In order to disrupt the primary processing of faces (and the intuitive accessibility of trait impressions), we asked participants to indicate initial decisions that were solely based on case-relevant information. They were then shown the entire case file, which also included facial photographs of the plaintiff and defendant, and they could still revise their sentencing decisions. As intended, the majority of participants (ca. 90%) did not change their initial sentences, which means that most sentences reflected decisions that were made while being ignorant of the plaintiff's and defendant's facial appearance. However, participants who decided to change their initial decisions overwhelmingly did so to bring their final decisions in line with facial stereotypes (e.g., by finding an untrustworthy-looking defendant guilty). The same pattern was observed for a second intervention in which participants were additionally prompted to make reflective decisions. Overall, this increased the influence of facial stereotypes.

[3] We also compared the rate of guilty verdicts in participants' initial sentencing decisions (i.e., when they were not exposed to facial photographs). Compared to the sequential condition (54.23%), participants in the sequential-and-reflection condition indicated slightly fewer guilty verdicts (51.67%), suggesting that instructing participants to reflect on their sentencing decisions slightly decreased their likelihood of indicating a guilty verdict. However, this difference was only marginally significant, β = −0.112, *SE* = 0.062, *z* = 1.82, *p* = .069, 95% CI [−0.246, 0.018], *OR* = 0.89.

7. Internal meta-analysis

To estimate the influence of facial trustworthiness on sentencing decisions more precisely, we calculated the meta-analytic effect across our three studies. We aggregated the data from all conditions that did not feature an intervention (the pretest and the control conditions from Study 1 and Study 2). This data set included almost 10000 sentencing decisions ($n = 967$ participants, 48.09% female, $M_{\text{age}} = 35.85$, $SD_{\text{age}} = 10.45$). We estimated a multilevel regression model with random intercepts and slopes per participant, per case, and per study. This revealed a positive effect of facial trustworthiness on sentencing decisions, $\beta = 0.284$, $SE = 0.052$, $z = 5.46$, $p < .001$, 95% CI [0.171, 0.398], $OR = 1.33$.[4] Defendants were more likely to be found guilty for the same transgression when they looked untrustworthy. The rate of guilty verdicts was 6.88 percentage points higher for untrustworthy-looking defendants (54.54% vs. 47.66%).

8. General discussion

The aim of the current investigation was to test the effectiveness of different interventions in reducing the influence of facial stereotypes on legal decision-making. We created a novel legal sentencing paradigm in which participants indicated verdicts for multiple small claims court cases and we manipulated the facial trustworthiness of plaintiffs and defendants. In line with previous studies showing that trait impressions from faces influence legal decision-making (Porter et al., 2010; Wilson & Rule, 2016), we found that defendants were more likely to be found guilty when they looked untrustworthy (vs. trustworthy). This effect was observed in all three studies. In our pretest, we also examined whether facial trustworthiness influences the fraction of damages that defendants were ordered to pay in case of a guilty verdict. Again, untrustworthy-looking defendants experienced less leniency as they were ordered to pay slightly more damages. Our results replicate previous findings by Zebrowitz and McDonald (1991) who found that babyfacedness—a facial feature that is correlated with perceived trustworthiness (Berry & Zebrowitz McArthur, 1986; Zebrowitz & Montepare, 1992) —predicted verdicts and awarded damages. Crucially, their findings were based on a large sample of real small claims cases, which suggests that the current results may generalize beyond our experimental design to real-world sentencing decisions.

We then tested the effectiveness of two debiasing techniques—educating decision-makers and changing the decision-making environment (Soll et al., 2014)—in reducing the influence of facial trustworthiness on verdicts. In Study 1, we attempted to reduce the influence of facial stereotypes by educating people about the poor diagnostic value of their trait impressions. Specifically, we (a) educated participants about the biasing influence of inaccurate facial stereotypes and (b) confronted them with the low diagnostic value of their own trait impressions. Although both manipulations succeeded in lowering beliefs that personality traits can be accurately inferred from a person's facial appearance, they did not reduce the effect of facial stereotypes on sentencing decisions. Bayesian analyses indicated strong support for the null hypothesis of no difference between the control condition and the intervention conditions. Thus, regardless of whether or not participants were given clear information about the low diagnostic value of their trait impressions from faces, sentencing decisions were influenced by the facial trustworthiness of defendants.

In Study 2, we attempted to reduce the influence of facial stereotypes by disrupting the intuitive accessibility of trait impressions. To this end, we provided information sequentially. First, participants saw only case-relevant information and indicated a preliminary sentence. Then, participants saw the entire case file (including facial photographs) and indicated their final sentence. As intended, only a minority of initial sentences (< 10%) were changed. However, sentence revisions were strongly driven by facial stereotypes, with most revised decisions reflecting a stereotype-congruent verdict (e.g., untrustworthy-looking defendants being found guilty). On average, this actually *increased* the influence of facial stereotypes on verdicts. A similar pattern was observed when participants were additionally prompted to make reflective decisions.

Together, our results highlight the persistent influence of facial stereotypes on decision-making. Previous studies have shown that people rely on trait impressions even when other, more diagnostic cues are available (Jaeger et al., 2019a; Olivola et al., 2018; Olivola & Todorov, 2010). In a similar vein, the present results demonstrate that effects of trait impressions on decision-making persist even when participants receive clear information about how inaccurate facial stereotypes are (Study 1) and even when participants have to expand additional cognitive effort to rely on facial stereotypes (Study 2). Across all interventions, we consistently found that untrustworthy-looking defendants were more likely to be found guilty than trustworthy-looking defendants.

8.1. Limitations and future directions

We acknowledge that our education intervention in Study 1 may not have been strong enough to reduce behavioral reliance on facial stereotypes. However, other studies employing similar manipulations successfully reduced lay beliefs and related behaviors (Chiu, Hong, & Dweck, 1997; Levy et al., 1998). For example, Levy et al. (1998) exposed participants to short scientific articles written for a lay audience to manipulate beliefs in the malleability of personality traits. The manipulation successfully influenced lay beliefs, but also the extent to which participants relied on stereotypes when judging different social groups. Regardless, our intervention only had a small effect on beliefs and more intensive debiasing trainings might be necessary to change behavior (for examples, see Devine, Forscher, & Austin, 2013; Sellier, Scopelliti, & Morewedge, 2019).

One question that remains unanswered is why a later presentation of photographs increased the influence of facial stereotypes. Studies on the role of fluency in cue ordering (Dimov & Link, 2017), anchoring effects (Tamir & Mitchell, 2012; Tversky & Kahneman, 1974), and primacy effects in impression formation (Asch, 1946) all highlight the strong influence of information that is processed *first*. However, other investigations found a disproportionate influence of information that is processed *last* (i.e., recency effects; Sullivan, 2018). For example, when evaluating faces that display a series of expressions, trait impressions are more strongly influenced by the expression that was displayed last (Fang, van Kleef, & Sauter, 2018). In a similar vein, participants might have attributed more importance to facial photographs because they were the only new information that was displayed after they indicated their preliminary verdicts. To participants, this may imply that this information is relevant for their decisions (Clark & Haviland, 1977). More research is needed to systematically explore how the order in which facial appearance and other cues are processed affects the influence of facial stereotypes.

We do not doubt that certain manipulations could diminish or eliminate the effect of facial trustworthiness on verdicts. For example, providing unambiguous, outcome-relevant information has been shown to reduce reliance on stereotypes (Dovidio & Gaertner, 2000; Rezlescu et al., 2012). However, such decisive information (e.g., clear evidence that the defendant committed the crime) is often not available in real life. In many situations, such as legal sentencing, personnel selection, or voting, people have to make consequential decisions based on limited, ambiguous, or contradicting information. We therefore focused on

[4] This effect was significantly larger than the effect of facial trustworthiness on initial verdicts in Study 2, where participants were not exposed to the face images (648 participants in the sequential and sequential-and-reflection conditions), $\beta = 0.177$, $SE = 0.068$, $z = 2.60$, $p = .009$, 95% CI [0.044, 0.310], $OR = 1.19$.

testing the effectiveness of different interventions in a decision-making environment that resembles these conditions.

In the present studies, we always compared situations in which the plaintiff was trustworthy-looking and the defendant was untrustworthy-looking or vice versa. Thus, our data do now show whether sentencing decisions are more strongly influenced by the perceived trustworthiness of plaintiffs or defendants, or whether there are interaction effects between both parties' perceived trustworthiness (Zebrowitz & McDonald, 1991). Moreover, if sentencing decisions are driven by the difference in perceived trustworthiness of the plaintiff and defendant, manipulating both parties' perceived trustworthiness simultaneously would increase its overall effect. This suggests that the effect of facial trustworthiness on sentencing decisions may be smaller under different circumstances, such as when a trustworthy-looking plaintiff is suing a slightly less trustworthy-looking defendant.

Finally, it should be noted that some of our participants may have been exposed to the face stimuli before, which were taken from a publicly available face database (Ma et al., 2015). This could have reduced the effect of facial trustworthiness of sentencing decisions, as previous work suggests that non-naïveté in participants leads to smaller effect sizes (Chandler, Paolacci, Peer, Mueller, & Ratliff, 2015; Rand & Kraft-Todd, 2014). For example, one can imagine that participants who were familiar with the faces paid less attention to them when making sentencing decisions. However, our data do not suggest that a large number of participants responded carelessly. Participants did not click through the cases as fast as possible—the median response time for verdicts was around 35 s—and only a fraction of participants (ca. 1%) indicated the same verdict on all trials. We also excluded data from all participants who failed an attention check at the end of the study.

Evidence for the biasing influence of trait impressions is well-documented and researchers have called for attempts to curb this bias (Olivola et al., 2014; Porter et al., 2010; Wilson & Rule, 2015). We took a first step in this direction but, ultimately, we were unsuccessful in reducing the influence of facial stereotypes. To stimulate further research in this area, we have made all materials needed to implement the legal sentencing task that was used here publicly available. This task allows for within-subject manipulations of facial appearance (or of other cues such as race or gender), which is statistically powerful and provides an indicator of reliance on facial stereotypes at the participant level. We hope that our results will motivate others to design and test other kinds of interventions.

9. Conclusion

The present research replicates prior findings that legal sentencing decisions are influenced by facial stereotypes. Participants consistently found untrustworthy-looking defendants guilty more often than trustworthy-looking defendants. We also sought to curb this bias by educating people about how inaccurate their trait impressions are and by disrupting the intuitive accessibility of trait impressions. Crucially, both attempts did not succeed in reducing the effect of facial trustworthiness on sentencing decisions. The present findings show that people persistently rely on facial stereotypes when making decisions and that this bias is difficult to mitigate.

Open practices

All data, materials, preregistrations, and analysis scripts are available at the Open Science Framework (https://osf.io/h4yf3/).

References

Alaei, R., & Rule, N. O. (2016). Accuracy of perceiving social attributes. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.). *The social psychology of perceiving others accurately* (pp. 125–142). Cambridge University Press. https://doi.org/10.1017/cbo9781316181959.006.

Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not so fast: The (not-quite-complete) dissociation between accuracy and confidence in thin-slice impressions. *Personality and Social Psychology Bulletin, 26*(2), 264–277. https://doi.org/10.1177/0146167209354519.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290.

Axt, J. R., Casola, G., & Nosek, B. A. (2018). Reducing social judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin, 45*(8), 1232–1251. https://doi.org/10.1177/0146167218814003.

Bainbridge, W. A., Isola, P., Blank, I., & Oliva, A. (2013). Establishing a database for studying human face photograph memory. In N. Miyake, D. Peebles, & R. P. Coopers (Eds.). *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1302–1307). Cognitive Science Society.

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition, 8*(4), 343–361.

Berry, D. S., & Zebrowitz McArthur, L. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin, 100*(1), 3–18. https://doi.org/10.1037/0033-2909.100.1.3.

Berry, D. S., & Zebrowitz-McArthur, L. A. (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin, 14*(1), 23–33.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General, 142*(1), 143–150. https://doi.org/10.1037/a0028930.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can we detect cooperators by looking at their face? *Current Directions in Psychological Science, 26*(3), 276–281. https://doi.org/10.1177/0963721417693352.

Bruce, V., & Young, A. (2012). *Face perception.* Psychology Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Univer. California Press.

Chan, M.-P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science, 26*(7), 1131–1139. https://doi.org/10.1177/0956797615585115.

Chiu, C., Hong, Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73*(1), 19–30. https://doi.org/10.1037/0022-3514.73.1.19.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.). *Discourse production and comprehension* (pp. 1–40). Ablex Publishing Corporation. https://doi.org/10.2307/1421524.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*(5), 784–803. https://doi.org/10.1037/0022-3514.90.5.784.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Devine, P. G., Forscher, P. S., & Austin, A. J. (2013). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003.

Dimov, C. M., & Link, D. (2017). Do people order cues by retrieval fluency when making probabilistic inferences? *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2002.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*(4), 315–319.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy. *Psychological Science, 17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x.

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports, 3*, 1047. https://doi.org/10.1038/srep01047.

Evans, A. M., & Krueger, J. I. (2016). Bounded prospection in dilemmas of trust and reciprocity. *Reviews of General Psychology, 20*(1), 17–28. https://doi.org/10.1007/s13398-014-0173-7.2.

Fang, X., van Kleef, G. A., & Sauter, D. A. (2018). Person perception from changing emotional expressions: Primacy, recency, or averaging effect? *Cognition and Emotion, 32*(8), 1597–1610. https://doi.org/10.1080/02699931.2018.1432476.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change in implicit bias. *Journal of Personality and Social Psychology, 117*(3), 522–559. https://doi.org/10.1037/pspa0000160.

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20*(5), 362–374. https://doi.org/10.1016/j.tics.2016.03.003.

Ghaffari, M., & Fiedler, S. (2018). The power of attention: Using eye gaze to predict other-regarding and moral choices. *Psychological Science, 29*(11), 1878–1889. https://doi.org/10.1177/0956797618799301.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior.* Oxford University Press.

Gigerenzer, G., Martignon, L., Hoffrage, U., Rieskamp, J., Czerlinski, J., & Goldstein, D. G.

(2008). One-reason decision making. In C. R. Plott, & V. L. Smith (Vol. Eds.), *Handbook of experimental economics*. *Vol. 1*. *Handbook of experimental economics* (pp. 1004–1017).

Graham, J. R., Harvey, C. R., & Puri, M. (2017). A corporate beauty contest. *Management Science, 63*(9), 3044–3056. https://doi.org/10.1287/mnsc.2016.2484.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498. https://doi.org/10.1111/2041-210X.12504.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4.

Greenwald, A. G., McGhee, D. E., Jordan, L. K., & 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.

Gunnell, J. J., & Ceci, S. J. (2010). When emotionality trumps reason: A study of individual processing style and juror bias. *Behavioral Sciences & the Law, 28*(2), 211–223. https://doi.org/10.1002/bsl.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*(6), 438–456. https://doi.org/10.1037/h0040736.

Hershberger, P. J., Markert, R. J., Part, H. M., Cohen, S. M., & Finger, W. W. (1997). Understanding and addressing cognitive bias in medical education. *Advances in Health Sciences Education, 1*, 221–226. https://doi.org/10.1007/BF00162919.

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019a). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General, 148*(6), 1008–1021. https://doi.org/10.1037/xge0000591.

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019b). Who judges a book by its cover? The prevalence, structure, and correlates of lay beliefs in physiognomy. https://psyarxiv.com/8dq4x/.

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1353–1361. https://doi.org/10.1037/a0027078.

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *Quarterly Journal of Experimental Psychology, 63*(11), 2273–2287. https://doi.org/10.1080/17470211003770912.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Joy-Gaba, J. A., Ho, A. K., ... Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Social Psychology, 143*(4), 1765–1785. https://doi.org/10.1037/a0036260.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421–1536. https://doi.org/10.1037/0022-3514.74.6.1421.

Ling, L., Luo, D., & She, G. (2019). Judging a book by its cover: The influence of physical attractiveness on the promotion of regional leaders. *Journal of Economic Behavior and Organization, 158*, 1–14. https://doi.org/10.1016/j.jebo.2019.01.005.

Lown, C. (1977). Legal approaches to juror stereotyping by physical characteristics. *Law and Human Behavior, 1*(1), 87–100. https://doi.org/10.1007/BF01044779.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5.

McKenzie, C. R., & Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research, 48*, S1–S13. https://doi.org/10.1509/jmkr.48.SPL.S1.

Mieth, L., Bell, R., & Buchner, A. (2016). Cognitive load does not affect the behavioral and cognitive foundations of social cooperation. *Frontiers in Psychology, 7*, 1–14. https://doi.org/10.3389/fpsyg.2016.01312.

Morewedge, C. K., & Kahneman, D. (2017). Associative processes in intuitive judgment. *Trends in Cognitive Sciences, 14*(10), 435–440. https://doi.org/10.1016/j.tics.2010.07.004.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights From the Behavioral and Brain Sciences, 2*(1), 129–140. https://doi.org/10.1177/2372732215600886.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-4.1. https://cran.r-project.org/package=BayesFactor.

Newman, I. R., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 43*(7), 1154–1170. https://doi.org/10.1037/xlm0000372.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007.

Olivola, C. Y., Tingley, D., & Todorov, A. (2018). Republican voters prefer candidates who have conservative-looking faces: New evidence from exit polls. *Political Psychology, 39*(5), 1157–1171. https://doi.org/10.1111/pops.12489.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*(2), 315–324. https://doi.org/10.1016/j.jesp.2009.12.002.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184–188. https://doi.org/10.1177/0963721414531598.

Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology, 74*, 8–23. https://doi.org/10.1016/j.jesp.2017.07.009.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 607–640. https://doi.org/10.1521/soco.2006.24.5.607.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*(6), 477–491. https://doi.org/10.1080/10683160902926141.

R Core Team (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Rand, D. G., & Kraft-Todd, G. T. (2014). Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience, 8*(September), 1–8. https://doi.org/10.3389/fnbeh.2014.00300.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS One, 7*(3), Article e34293. https://doi.org/10.1371/journal.pone.0034293.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12*(1), 94–99. https://doi.org/10.1111/1467-9280.00317.

Rule, N. O., Bjornsdottir, R. T., Tskhay, K. O., & Ambady, N. (2016). Subtle perceptions of male sexual orientation influence occupational opportunities. *Journal of Applied Psychology, 101*(12), 1687–1704. https://doi.org/10.1037/apl0000148.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*(3), 409–426. https://doi.org/10.1037/a0031050.

Rule, N. O., Tskhay, K. O., Freeman, J. B., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology, 44*(6), 529–535. https://doi.org/10.1002/ejsp.2043.

Sellier, A.-L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training transfers to improve decision making in the field. *Psychological Science, 30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429.

Shah, A. K. (2007). Easy does it: The role of fluency in cue weighting. *Judgment and Decision making, 2*(6), 371–379. https://doi.org/10.1037/e722852011-015.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207.

Shevlin, M., Walker, S., Davies, M. N. O., Banyard, P., & Lewis, C. A. (2003). Can you judge a book by its cover? Evidence of self-stranger agreement on personality at zero acquaintance. *Personality and Individual Differences, 35*(6), 1373–1383. https://doi.org/10.1016/S0191-8869(02)00356-2.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135*(3), 409–428. https://doi.org/10.1037/0096-3445.135.3.409.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. https://doi.org/10.2307/1884852.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2014). A user's guide to debiasing. In G. Wu, & G. Keren (Eds.). *Handbook of judgment and decision-making.* Wiley.

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General, 141*(4), 715–727. https://doi.org/10.1037/a0027950.

Stoker, J. I., Garretsen, H., & Spreeuwers, L. J. (2016). The facial appearance of CEOs: Faces signal selection but not performance. *PLoS One, 11*(7), Article e0159950. https://doi.org/10.1371/journal.pone.0159950.

Sullivan, J. (2018). The primacy effect in impression formation: Some replications and extensions. *Social Psychological and Personality Science, 10*(4), 1–8. https://doi.org/10.1177/1948550618771003.

Suzuki, A., Tsukamoto, S., & Takahashi, Y. (2017). Faces tell everything in a just and biologically determined world. *Social Psychological and Personality Science, 10*(1), 62–72. https://doi.org/10.1177/1948550617734616.

Tamir, D. I., & Mitchell, J. P. (2012). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General, 142*(1), 151–162. https://doi.org/10.1037/a0028232.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657–665. https://doi.org/10.1080/13506280500410949.

Tiddeman, B., Burt, M., & Perrett, D. I. (2001). Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications, IEEE, 21*(5), 42–50.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions.* Princeton University Press.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*(1), 72–86. https://doi.org/10.1177/1088868312461308.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science, 7*(4), 331–338. https://doi.org/10.1177/1948550615624142.

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

Yu, M., Saleem, M., & Gonzalez, C. (2014). Developing trust: First impressions and experience. *Journal of Economic Psychology, 43*, 16–29. https://doi.org/10.1016/j.joep.2014.04.004.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*(6), 603–623. https://doi.org/10.1007/BF01065855.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced males and females across the lifespan. *Developmental Psychology, 28*(6), 1143–1152.