It's not all Black and White: implicit racial bias extends to other races as well

Marvin R. Maechler
Dartmouth College
Vassiki S. Chauhan
Dartmouth College
Natalia A. McLaren
Dartmouth College
Samantha M. Norton
Dartmouth College
Jae S. Hong
Dartmouth College
M. Ida Gobbini (mariaida.gobbini@unibo.it)
University of Bologna

Article

Keywords:

Posted Date: March 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1432665/v1

Abstract

Here we investigated the role of racial attitudes in trait inferences from facial appearance. In two separate experiments we measured implicit racial attitudes using the race *Implicit Association Test* (IAT), as well as judgments of several traits after brief exposure to strangers' faces. Results from the first experiment (n=67) suggested that judgement of attractiveness is biased by implicit skin tone preference. The second experiment (n=104) extended this finding by testing four additional personality traits (attractiveness, aggression, competence, likeability, and trustworthiness). In both experiments, implicit racial attitudes predicted differential trait ratings of White and Black faces. Additionally, we found that implicit bias measured with the IAT predicted bias in trait inferences between White and Asian, but not between Black and Asian faces. These findings suggest that an implicit racial attitude could be motivated by a more general tendency to judge White faces differently from non-White faces.

Statement Of Relevance

First impressions are ubiquitous and spontaneous. We found that implicit racial attitudes towards White versus Black faces have a diffuse effect on first impressions that generalizes to East Asian faces. Our results indicate that racial bias, as measured by the Implicit Association Test, is not simply Black and White, but reveals a more general attitude towards racial disparities.

Introduction

Seeing a face for the first time elicits spontaneous inferences about the personality traits of that person. The notion that facial appearance is related to inner character is common and has a long tradition which was formally conceptualized in a book on physiognomy by Lavater (1775–1778; Hassin & Trope, 2000). More recently, research has shown not only that trait inferences are pervasive but also that there is a surprising consistency in adults' trait impressions based on facial appearance (e.g., Carré, McCormick, & Mondloch, 2009; Zebrowitz & Rhodes, 2002, Oosterhof & Todorov, 2008). While the ability to make quick decisions given limited information is an adaptive behavioral skill, making decisions about who a person is, given how they look, runs the risk of amplifying implicit biases and racial stereotypes to which all humans are susceptible (Marini & Banaji, 2020).

Implicit biases can be elicited by different types of social cues such as socioeconomic status, or cultural, national, racial, ethnic, and gender identity (Kaufmann et al., 2017; Oh, Shafir, Todorov, 2020; Olivola & Todorov, 2010; Zebrowitz & Montepare, 2010). Racially based stereotypes, feelings, beliefs, and notions can lead to negative judgements of individuals of other races and can have a profound impact on interaction between individuals of different races (Pennington et al., 2016). Dovidio and Gaertner (1986), in their theory of Aversive Racism, posited that stereotypical qualities of Black people and other racial minorities affected how White Americans reacted around them even when maintaining that discrimination is wrong. Despite the strong subjective impression that one might be racially unbiased, it has been shown that subliminal biases appear not in self-report but, for instance, in measures of association strength between skin color (which informs, but does not fully explain, perceived race and ethnicity) (Uhlmann et al., 2002) and affectively valenced words (Greenwald, Ghee & Schwartz, 1998). Implicit racial attitudes have subtle but measurable influences on the judgement of trustworthiness in strangers' faces (Stanley, Sokol-Hessner, Banaji, & Phelps, 2011). Such implicit tests allow us to assess biases that people are either unaware of or might not want to disclose.

One established measurement device for these biases, the *Implicit Association Test* (Greenwald, Ghee & Schwartz, 1998; Lane, Kang & Banaji, 2007), has been used extensively in social psychology research. Several meta-analyses show that this measure is reliably replicable, although there are debates in the field about the extent to which it can predict biased behavior (Greenwald, Poehlman, Uhlmann & Banaji, 2009; Blanton, Jaccard, Klick, Mellers, Mitchell, & Tetlock, 2009). A meta-analysis of the race IAT has found that it explains on average 5% of variance in racially biased behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). The race IAT is suitable to reveal a bias that exists within a population, but it is not well suited to predicting behavior on an individual basis (Blanton, Jaccard, Klick, Mellers, Mitchell, & Tetlock, 2009; Hehman, Calanchini, Flake & Leitner, 2019). Therefore, it should not be interpreted as a diagnostic tool for measuring racial bias on the participant level. However, there is some evidence that the IAT could be diagnostic of biases at the population level (Hehman, Calanchini, Flake & Leitner, 2019). This metric, as the name clearly suggests, is simply an approximate measure of implicit bias. Because of its extensive use in research in social psychology, we have adopted the IAT in the present study to measure a racial bias toward White and Black faces (Gawronski, 2002; Vianello & Bar-Anan, 2020).

The aim of the present study was to investigate whether there is a reliable relationship between implicit racial biases and the way we make snap judgments about new faces. In contrast to previous studies, we focused on judgments of attractiveness as well as inferences about several different personality traits based on facial appearance. We tested trait inferences for Black faces, White faces and faces from a third race: East Asian faces. We found that implicit racial biases predict biases in ratings of attractiveness and inferred personality traits, with implicit preference for Black faces predicting higher attractiveness ratings for Black faces compared to White faces (and vice versa). We were also able to predict how a participant would rate attractiveness and some personality traits of East Asian faces relative to White faces based on skin tone preference as measured with the IAT. Stronger implicit preferences for White faces predicted more positive ratings of White faces relative to East Asian faces. The race IAT did not predict how a participant would judge East Asian faces relative to Black faces.

Method Experiment 1

Participants

67 Dartmouth undergraduate students participated in the first experiment (48 women and 19 men, average age = 19.37, *sd* = 1.11). All participants completed the experiment, and no participant was excluded from the sample. Of our participants, roughly half (50.7%) identified their own race as White (see the supplementary information for a detailed breakdown of participants' self-identified race

identities). Participants had normal or corrected to normal vision. The study protocol was approved by the Institutional Review Board at Dartmouth College (Protocol 29780) and was done in accordance with the relevant guidelines and regulations including the declaration of Helsinki. All participants gave informed consent and were reimbursed with course credit in an introductory psychology class. The specific sample size was chosen to achieve 80% power assuming a moderate correlation ($r = .3$) between implicit and explicit bias based on a pilot sample, which was collected as part of the culminating experience of a class project on research design. Said pilot sample was not reported here.

Procedure

The experimental protocol consisted of two parts: first, participants rated the attractiveness of briefly presented frontal face images, then they completed the race IAT (Greenwald, McGhee & Schwartz, 1998). Data collection took no longer than 25 minutes per participant. The participants were not told that the true purpose of the experiment was to uncover how potential racial bias affects perception of attractiveness of faces of strangers, since this kind of *stereotype threat* has been shown to impact the results of the IAT (Frantz, Cuddy, Burnett, Ray & Hart, 2004). Instead, we told participants that we were interested in their subjective judgement of the attractiveness of faces, and that there were no wrong answers.

We collected attractiveness ratings for 48 stimuli: East Asian, Black, or White faces (8 images per race-gender-combination – 48 images total). With the purpose of ensuring only first impressions from faces, we presented each of them separately and only once. Participants self-initiated each trial with a button press, after which a face image appeared for 500ms. After its disappearance, participants saw a prompt to rate the face. Participants judged the attractiveness of each face immediately after the image disappeared by adjusting a slider with the computer mouse. The slider moved a black marker from left (very unattractive) to right (very attractive) along a white line spanning the whole screen. Underneath the white line there were equally spaced numbers ranging from zero to ten (0 - very unattractive, 10 - very attractive). These numbers served as a visual guideline to the participant, while we used the more accurate exact position along the slider (i.e., the pixel where the participant put the rating) in our analyses, which allowed for a higher resolution of attractiveness ratings (Fig. 1).

Exposure to each face for half of a second ensured that participants would only form a first impression about the face. Previous research has shown that seeing an unfamiliar face for a very brief period (e.g., 100ms) is sufficient to obtain a reliable first impression along different trait dimensions like attractiveness (Willis & Todorov, 2006). Similarly, it has been shown previously that observers extract demographic information like race from viewing a face for only 100ms (Colombatto, Uddenberg & Scholl, 2021).

Next, we administered a common version of the race IAT and specifically followed recommendations from several methodological reviews on its use (Nosek, Greenwald & Banaji, 2007; Greenwald, Nosek & Banaji, 2003). A newer review with more recommendations on using the IAT (Greenwald et al., 2021) was not available when we collected the data in 2019/2020, however, our procedures were in line with the recommendations outlined therein. This test measures the association strength between White/Black faces and negative/positive words, respectively. Participants are required to associate a set of faces (dark-skinned or light-skinned) with a set of words (positive or negative). If they are faster and more accurate when they associate White faces with negative words and Black faces with positive words than with the opposite pairings, they are said to have a preference for dark-skinned faces (and vice versa in the opposite case). In other words, this test measures one's implicit readiness to assign biased labels to faces with different skin tones and is sensitive to the direction of that bias.

Stimuli

Face images

We used face images from the Multi-Racial Mega-Resolution (MR2) database (Strohminger et al., 2016), which we downloaded from the open science framework (https://osf.io/skbq2/). Out of the 74 face images in this database, we selected eight images per race (Black, White, and East Asian) and gender group (male or female). We used the attractiveness ratings provided with the database to ensure that they were on average the same for each race (see Fig. 2 for average attractiveness ratings per race group).

This database was chosen for its high degree of realism (due to the high resolution of the images) and prior validation for attractiveness ratings. Furthermore, faces from the MR2 database show a low degree of variability in irrelevant features for the purpose of our experiment such as variable viewing angles and different emotions. The viewing distance, lighting, positioning, hair, and make-up are all highly controlled for all the images, which allowed us to rule out the influence of these factors on the ratings obtained in the first part of the experiment (Strohminger et al, 2016).

Implicit Association Test

The *Implicit Association Test* was first introduced by Greenwald and colleagues (Greenwald, Ghee & Schwartz, 1998). Black and White face images for the race IAT were downloaded from a website set up by, among others, one of its original authors (https://www.projectimplicit.net/resources/study-materials/; Nosek et al., 2007b). The positive and negative words were selected such that each affectively positive word (Ambition, Delight, Happy, Kindness, Triumph) would have an affectively negative counterpart with the same initial letter (Abuse, Depression, Hatred, Killer, Trouble). We decided to opt out of running the test on the website provided by some of the original authors and instead chose to write the stimulus code in Matlab (The MathWorks, Natick, MA, USA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Running the experiment offline rather than through a website allowed us to obtain more exact reaction times independent of network connection strength or the kind of computer a participant might have used. This also made it possible to calculate the exact value of *D*, the most validated measure of implicit bias strength (Greenwald, Nosek & Banaji, 2003; Nosek, Greenwald & Banaji, 2007).

The bias is calculated as follows. In the IAT, the observer has to sort categories: Black and White faces and positive and negative words using button presses with the left and right hand. On some blocks the same hand is used to classify one race and one valence, this mapping is switched on other blocks of the experiment. For example, the instructions of one block of trials are to press the left button for Black faces and positive words, while pressing the right button for White faces and negative words, then switch the choice of the button press after some trials (i.e, the left button for White faces and positive words, etc). The average response latency of a block of trials with one such mapping is then subtracted from the average response latency of an equally long block of trials with the opposite mapping. This is then normalized using the pooled standard deviation of the responses. If someone is faster at associating one race with the positively valenced words, they are said to prefer faces with that skin tone. The resulting measure of bias is called *D* and the optimal way to calculate it for specific applications of the IAT has been previously described in detail (see Greenwald, Nosek & Banaji, 2003).

Transparency and Openness

Datasets analyzed in this study are publicly available in the Open Science Framework repository (https://osf.io/vycnq/). We report how we determined our sample size (see the supplementary material for details on the power analyses), all data exclusions, all manipulations, and all measures in the study. Stimuli were presented using MATLAB (The MathWorks, Natick, MA, USA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), we also used MATLAB to aggregate the data prior to analysis. Data were analyzed using R, version 4.0.2 (R Core Team, 2020) and ggplot2, version 3.3.5 (Wickham, 2016). Databases and websites from which we acquired the study materials have been cited and linked to where appropriate. This study's design and the analyses were not pre-registered.

Results Experiment 1

Explicit skin tone preference values were calculated for each participant by z-scoring their attractiveness ratings from all faces, and subsequently subtracting the average rating of faces with one skin color from the average rating of faces with another skin color. This way, a score of zero corresponds to a neutral disposition, while positive numbers indicate a preference for one race, and negative numbers indicate a preference for the other race. The unit of this score is the number of standard deviations between the two means. Bias scores between Black and White faces (Fig. 3A) were calculated such that positive numbers indicate a preference for Black faces. In the analysis of bias between Asian and White faces (Fig. 3B) positive numbers indicate a preference for Asian faces. Finally, when looking at bias between Asian and Black faces (Fig. 3C) positive numbers indicated a preference for Black faces.

The implicit skin tone preference score *D* was calculated from the reaction times and errors in the IAT as described in a methodological review on its scoring by one of its original authors (Greenwald, Nosek & Banaji, 2003). *D* is bounded between − 2 and 2, where positive numbers indicate a dark skin tone preference. This number is derived from the difference of reaction time distributions in congruent and incongruent blocks of the IAT and its unit corresponds roughly to one standard deviation of difference.

Implicit bias as measured with the IAT was significantly predictive of explicit skin tone bias in the attractiveness ratings when looking at bias between Black and White faces ($r(65) = .307$, $p = .012$, $R^2 = .094$) and also when looking at bias between Asian and White faces ($r(65) = .446$, $p < .001$, $R^2 = .199$). However, there was no significant relationship between race IAT measures and bias between Asian and Black faces ($r(65)=-.141$, $p = .255$, $R^2 = .020$). See Fig. 3 for details. These relationships among bias values were robust to the exclusion of data from participants who identified their race as Asian, for whom the Asian category would have been a kind of in-group, although this obviously reduced statistical power. The same is true for exclusion of participants who identified as White.

Intermediate Discussion

"Beauty lies in the eye of the beholder" and, according to our data, it is influenced by the beholder's implicit racial attitude. Our results show that racial bias in attractiveness judgements is predicted by the race IAT (Greenwald, McGhee & Schwartz, 1998). Interestingly, our data suggest that an implicit bias between Black and White faces also extends to the judgment of attractiveness of faces from a race that was not featured in the IAT, East Asians, suggesting that those participants who have a skin tone preference for Black over White faces also have a more positive attitude toward East Asians over White faces (and vice versa).

Attractiveness is related to the appearance of a face, however other more complex judgements such as personality traits are made based on facial appearance as well (e.g., trustworthiness or competence; Willis & Todorov, 2006). In order to see whether implicit skin tone preference influences judgement only of the structural features of a face such as attractiveness or also affects judgments of personality traits, we collected a further dataset. We measured implicit bias for Black and White faces using the IAT (as in experiment 1) and explicit assessment of personality traits of Black, White, and East Asian faces.

Methods Experiment 2

Participants

In our second experiment, 104 undergraduate students from the Dartmouth community, who had not participated in experiment 1 and were naive to the purpose of the experiment, were recruited (74 women and 30 men, mean age 18.6 years *sd* = 0.9 years). Four participants were excluded, two due to technical errors and two who withdrew from the experiment. Of all participants, 62.5% identified their own race as White (see the supplementary information for a detailed breakdown of participants' self-identified race identities). Participants' vision was either normal or corrected to normal. As in experiment 1, participants were not told that the true purpose of the experiment was to uncover how potential racial bias affects judgments based on facial appearance. We told participants that we were interested in their subjective judgement of the faces presented during the experiment and that there were no wrong answers. The study protocol was approved by the Institutional Review Board at Dartmouth College (Protocol 29780) and was done in accordance with the relevant guidelines and regulations including the declaration of

Helsinki. All participants gave informed consent and were reimbursed with course credit for an introductory psychology class.

Due to the outbreak of the COVID-19 pandemic, we fell short in collecting our target sample size for achieving 90% power with our experimental design (we collected data from 104 participants while the target was 107. We based this target n on the assumption that r would be 0.307 and p = 0.025, correcting for multiple comparisons).

Procedure and Stimuli

The stimuli were the same as in experiment 1. In addition to attractiveness, we asked participants to rate each face along four more traits: aggression, trustworthiness, competence, and likeability. These are the same traits that were used by Willis and Todorov (2006), who revealed two major factors comprising first impressions. To ensure first impression judgment, faces were presented only once, and all five traits were rated simultaneously and immediately after the presentation of the face image. The experimental session started with five practice trials using faces that were not presented for the real experiment. The practice trials were not included in the final analysis. The race IAT session followed the rating of faces, and it had the same structure as described in Experiment 1.

Following Willis and Todorov (2006) we performed a principal components analysis on the aggregate scores from each face to reveal a lower dimensional space that describes the factor structure of these trait ratings. We calculated bias scores for these factors for each participant. First, we calculated the dot product of factor loadings and trait ratings to determine how a participant had rated a face along the factor. Then we subtracted the average factor score of the White faces from the average factor score of the Black faces, creating bias scores for each factor and participant. Biases between the other races were calculated in the same fashion. We used these factor bias scores to find correlations with implicit bias, as measured with the IAT.

Results Experiment 2

Principal-components analysis on the aggregate trait ratings of all images identified two orthogonal factors with eigenvalues greater than one. The first factor had its largest loadings on aggression, competence, and trustworthiness and accounted for 47.6% of the variance, whereas the second factor accounted for 41.2% of the variance and had its largest loadings on attractiveness and likeability. These factors are highly similar to those found by Willis and Todorov (2006) with the exception that they found that competence loaded more on the second factor than the first. See table 1 for details.

These factor biases for Black versus White faces were both significantly correlated with implicit bias (factor 1: $r(98) = .200$, $p = .046$, $R^2 = .04$; factor 2: $r(98) = .227$, $p = .023$, $R^2 = .052$). Biases in these factors for East Asian versus White faces were also correlated with implicit bias (factor 1: $r(98) = .226$, $p = .024$, $R^2 = .051$; factor 2: $r(98) = .214$, $p = .033$, $R^2 = .046$). We found that the IAT did not predict bias for Black versus East Asian faces on either of these factors (factor 1: $r(98) = .029$, $p = .774$, $R^2 < .001$; factor 2: $r(98) = .053$, $p = .597$, $R^2 = .003$). See Fig. 4 for a visualization of these relationships between biases.

These results show that the relationship of implicit bias to bias in judgements about strangers' faces extends to ratings of more complex personal traits such as trustworthiness, competence, aggressiveness, and likeability. The race IAT was again predictive of a more general White versus non-White bias. These relationships among bias values were robust to the exclusion of data from participants who identified their race as Asian, for whom the Asian category would have been a kind of in-group, although this obviously reduced statistical power. The same is true for exclusion of participants who identified as White.

Table 1. *Factors of trait judgements and their loadings*

	Present experiment		**Willis & Todorov (2006)**	
Trait	Factor 1	Factor 2	Factor 1	Factor 2
Aggression	− .96	− .07	− .96	− .01
Attractiveness	.07	.94	.00	.84
Competence	.84	.29	.06	.91
Likeability	.42	.85	.33	.79
Trustworthiness	.76	.60	.61	.61

Note. Principal-components analysis was done on the aggregated trait ratings for each face image. Only factors with eigenvalues greater than 1 were extracted. Compare with Willis and Todorov (2006) who found essentially the same factors, with the only notable difference in the loading of competence judgements.

Discussion

Here, we investigated the role of implicit racial bias in trait inferences about faces of strangers. Our data showed that implicit racial attitudes for Black versus White faces, as measured by the race IAT (Greenwald, McGhee & Schwartz, 1998), predict biases in judgments one makes about attractiveness and personality traits of briefly presented novel faces. Furthermore, the Black/White preference measured with the IAT extended to biases in assessment of faces from a third racial group (East Asian) with a preference for White over Black faces extending to a preference for White over Asian faces. It has already been demonstrated that implicit racial attitude predicts biased trustworthiness judgments (Stanley et al., 2011). We extended previous work by adding further personality traits, namely attractiveness, aggressiveness, competence, and likeability, and testing biases in judgments of faces from another race. Our results also suggest that racial bias against Black faces (or for White faces) may be a more diffuse trait that affects biased judgments of faces from other races as well.

In our second experiment, participants rated every face along five dimensions. PCA analysis revealed two factors that roughly correspond to competence/trustworthiness/aggression (with a negative weight for aggression) and attractiveness/likeability. These two factors of first impressions are in agreement with what has been reported previously by others (Willis & Todorov, 2006). Our results show that judgements along both of these factors are influenced by racial bias, suggesting that the race IAT measures a broad bias that affects most aspects of forming a first impression.

In addition to the race categories included in the race IAT, we acquired trait ratings for East Asian faces, which allowed us to test whether the race IAT predicts biases toward races that are not featured in the IAT. Our data suggest that the race IAT measures are not driven by a strict preference for either White or Black faces. Our data show that the IAT predicts the degree to which White faces are judged differently from Black and Asian faces. However, there seems to be no systematic relationship between race IAT scores and bias between Black and Asian face ratings. This might suggest that the race IAT measures something akin to a bias between White and non-White faces, rather than between White and Black faces. Our results are valid for the sample of participants – American college students – and generalization to other populations will require further research.

Our results constitute an example of predicting biased explicit behavior from IAT scores. It should, however, be noted that the IAT is not considered a diagnostic tool for bias in individuals, given that race IAT scores explain on average around 5% of variance in racially biased behavior (Blanton et al., 2009; Greenwald, et al., 2009). In the case of our experiment 1, the race IAT explained almost twice as much variance of racial biases in first impressions. In experiment 2, the race IAT explained variance to a lesser degree in comparison to the data reported in experiment 1 but was close to the average reported in previous literature. The difference between experiment 1 and 2 could be due to the additional cognitive load associated with rating faces across five different dimensions. Alternatively, participants might have been more careful about providing ratings when the demand was in considering personality traits beyond facial attractiveness.

In summary, implicit racial bias influences how we judge and infer personality traits from briefly viewed faces of strangers. Our results show that race IAT measures of bias between Black and White skin tones extend to other races, suggesting that the bias revealed by the IAT may be a more general bias distinguishing White faces from those of other races. Biased first impressions based on race may have consequences even for the behavior of individuals who do not harbor explicit racial attitudes. Future research could investigate further whether such generalized bias extends to other groups, as well, such as women, members of the LGBT community, and people from different social backgrounds (Ofosu et al., 2019).

Declarations

Data availability

The datasets generated and analyzed during the current study are available in the Open Science Framework repository, https://osf.io/vycnq/

Author contributions

MRM, MIG, NM, SN, and JH developed the study concept. MRM programmed the experiment code. NM, SN, and JH collected the data. MRM analyzed the data. MRM drafted the paper and VC, NM, and MIG provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

The experiments reported in the present paper were part of a project discussed by undergraduate students NM, SN, and JH under the supervision of MIG and MRM during a class on research methods. We thank Jim Haxby for access to his testing equipment. We thank Jim Haxby and Carlo Cipolli for helpful discussions. MIG was supported by NSF grant #1835200.

References

1. Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of applied Psychology*, 94(3), 567.
2. Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433-436.
3. Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological science*, 20(10), 1194-1198.
4. Colombatto, C., Uddenberg, S., & Scholl, B. J. (2021). The Efficiency of Demography in Face Perception. *Attention, Perception, & Psychophysics*, *83*(8), 3104-3117.
5. Dovidio, J. F., & Gaertner, S. L. (1986). Prejudice, discrimination, and racism. *Academic Press.*
6. Frantz, C. M., Cuddy, A. J., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, 30(12), 1611-1624.
7. Gawronski, B. (2002). What does the implicit association test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental psychology*, *49*(3), 171.
8. Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.
9. Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology*, 85(2), 197.

10. Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1), 17.
11. Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friese, M., ... & Wiers, R. W. (2021). Best research practices for using the Implicit Association Test. *Behavior research methods*
12. Hassin, R., & Trope, Y. (2000). Facing faces: studies on the cognitive aspects of physiognomy. *Journal of personality and social psychology*, 78(5), 837.
13. Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of experimental psychology: General*, *148*(6), 1022.
14. Kaufmann, M. C., Krings, F., Zebrowitz, L. A., & Sczesny, S. (2017). Age bias in selection decisions: The role of facial appearance and fitness impressions. *Frontiers in psychology*, *8*, 2065.
15. Lane, K., Kang, J., & Banaji, M. R. (2007). Implicit Social Cognition and Law. *Annual Review of Law & Social Science*, 3, 07-37.
16. Marini, M., & Banaji M. R. (2020). An implicit gender sex-science association in the general population and STEM faculty. *Journal of General Psychology*, 1-28. doi: 10.1080/00221309.2020.1853502.
17. Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007a). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior* (pp. 265-292). New York: Psychology Press.
18. Nosek, B.A., Smyth, F.L., Hansen, J.J., Devos, T., Lindner, N.M., Ratliff (Ranganath), K.A., Smith, C.T., Olson, K.R., Chugh, D., Greenwald, A.G. and Banaji, M.R., (2007b). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1), pp.36-88.
19. Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences*, *116*(18), 8846-8851.
20. Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature human behaviour*, *4*(3), 287-293.
21. Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of nonverbal behavior*, 34(2), 83-110.
22. Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087-11092.
23. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10, 437-442.
24. Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PloS one*, 11(1), e0146487.

25. R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/
26. Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, 108(19), 7710-7715.
27. Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior research methods*, 48(3), 1197-1204.
28. Sutherland, C. A., Burton, N. S., Wilmer, J. B., Blokland, G. A., Germine, L., Palermo, R., ... & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences*, 117(19), 10218-10224.
29. Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A. G., & Swanson, J. (2002). Subgroup prejudice based on skin color among Hispanics in the United States and Latin America. *Social Cognition*, *20*(3), 198-226.
30. Vianello M, Bar-Anan Y. (2021). Can the Implicit Association Test Measure Automatic Judgment? The Validation Continues*. Perspect Psychol Sci*. 16:415-421.
31. Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer International Publishing.
32. 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.
33. Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and personality psychology compass*, 2(3), 1497-1517.
34. Zebrowitz, L. A., & Rhodes, G. (2002). Nature let a hundred flowers bloom: The multiple ways and wherefores of attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 261–293). Ablex Publishing.

Figures

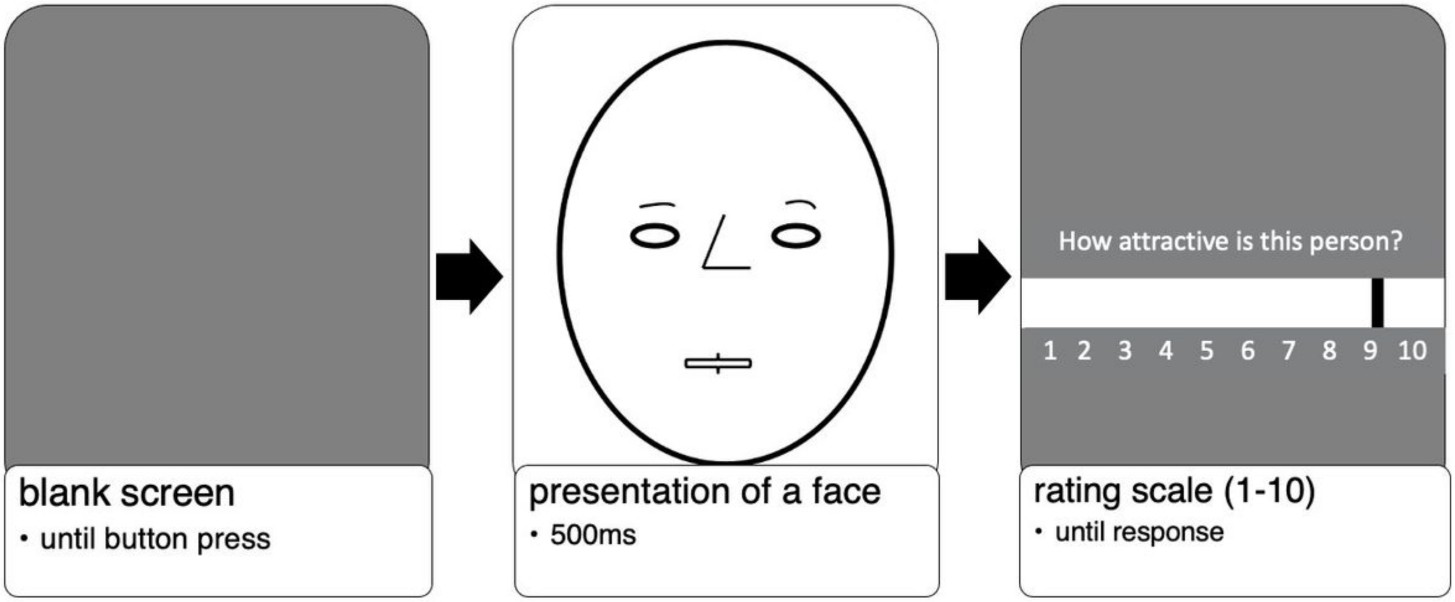

Figure 1

Example of the experimental paradigm for explicit ratings of faces (schematic). Participants initiate the trial and see a face for 500ms. Then, they use a sliding scale to give a rating. The “face image” in the second panel is merely a sketch, whereas the actual stimuli were taken from the MR2 database (Strohminger et al., 2016).

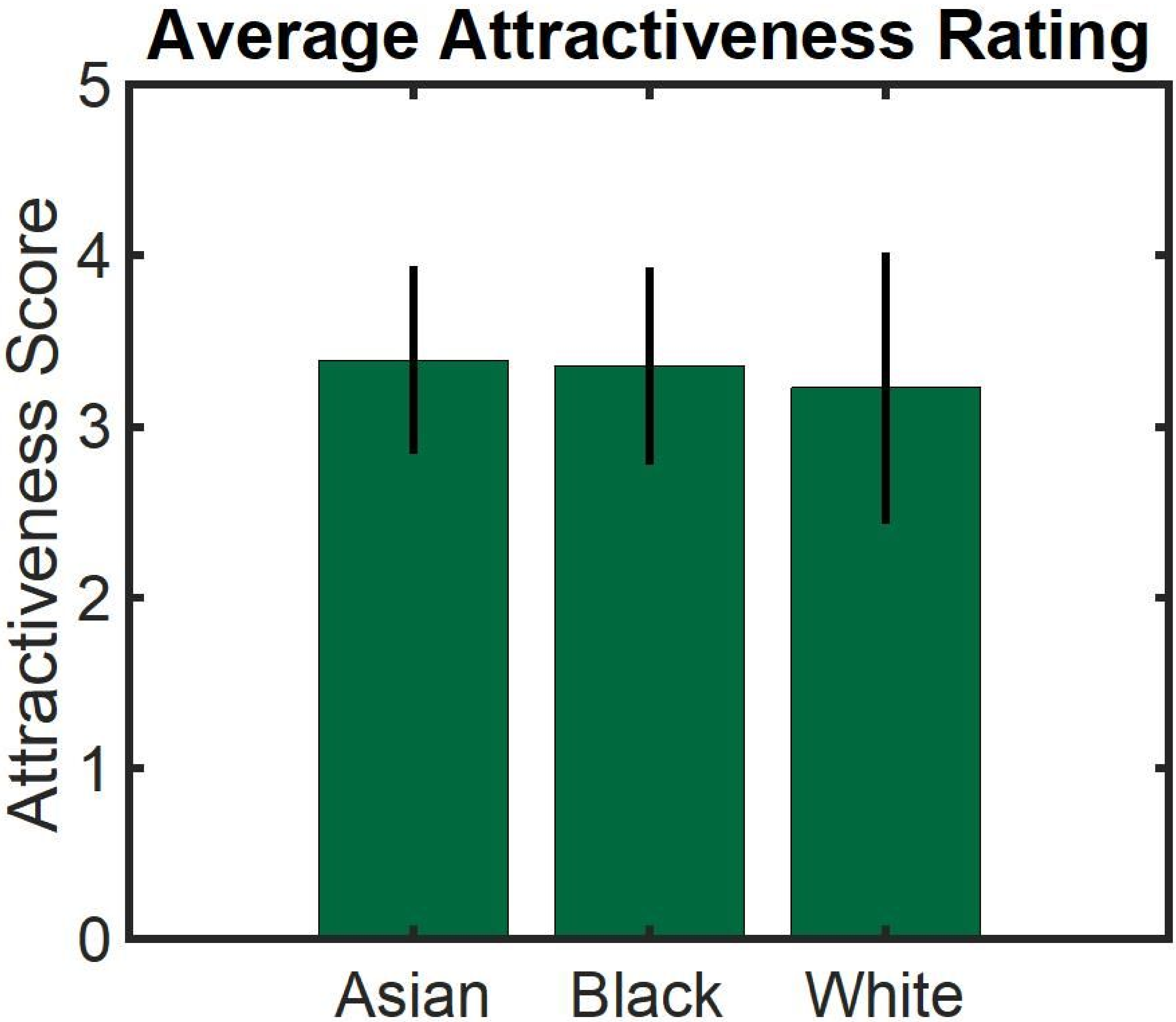

Figure 2

Average attractiveness rating of the images used in the first impression ratings. Bars show the average for each group of 16 images (8 male, 8 female) per race based on the validations published with the MR2 database (Strohminger et al., 2016). Error bars indicate one standard deviation from the mean.

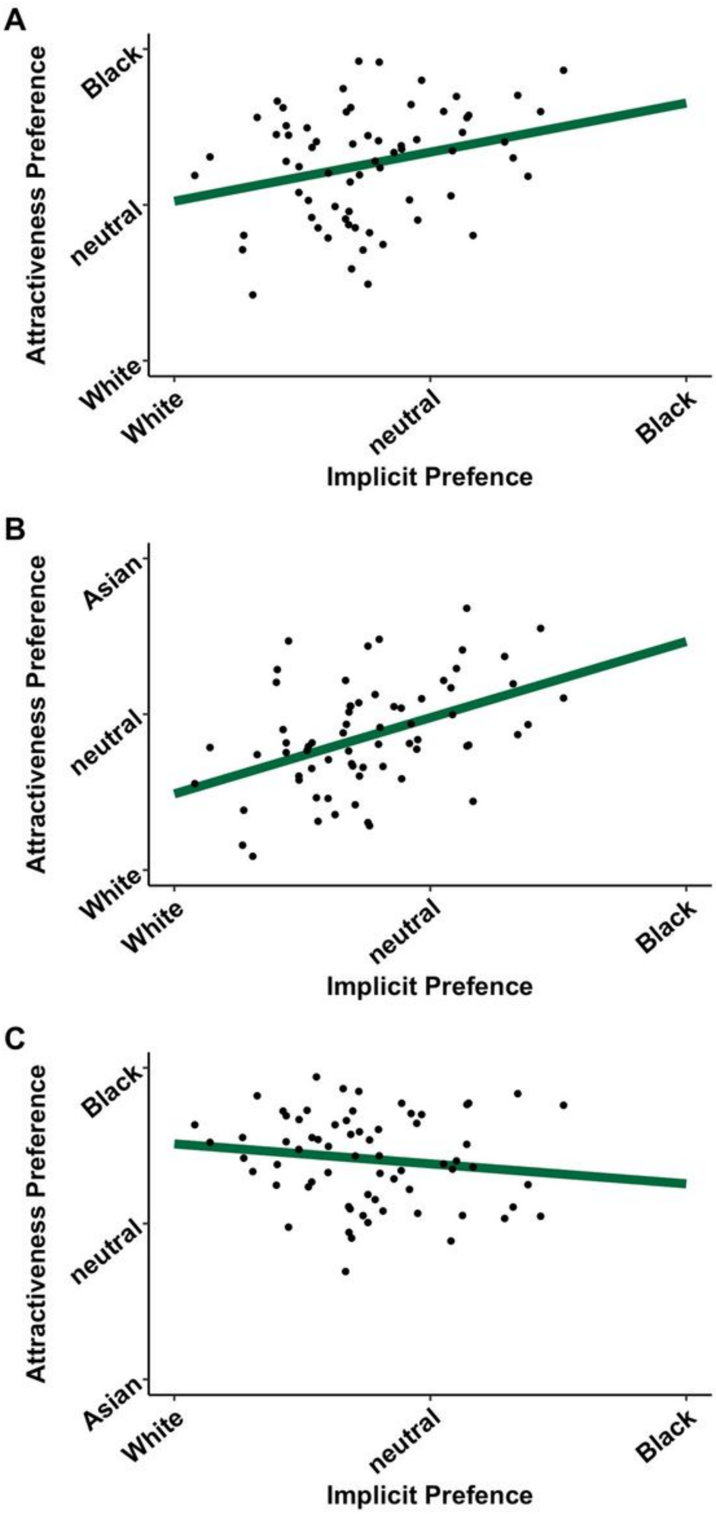

Figure 3

Implicit bias (x-axes) is significantly predictive of biases in attractiveness judgements (y-axes), between Black and White faces (A) and between Asian and White faces (B) but not for biases between Asian and Black faces (C). All axes are labelled to indicate which preference positive and negative scores indicate. Each black dot represents one participant, the line represents the best linear fit of implicit to explicit bias scores.

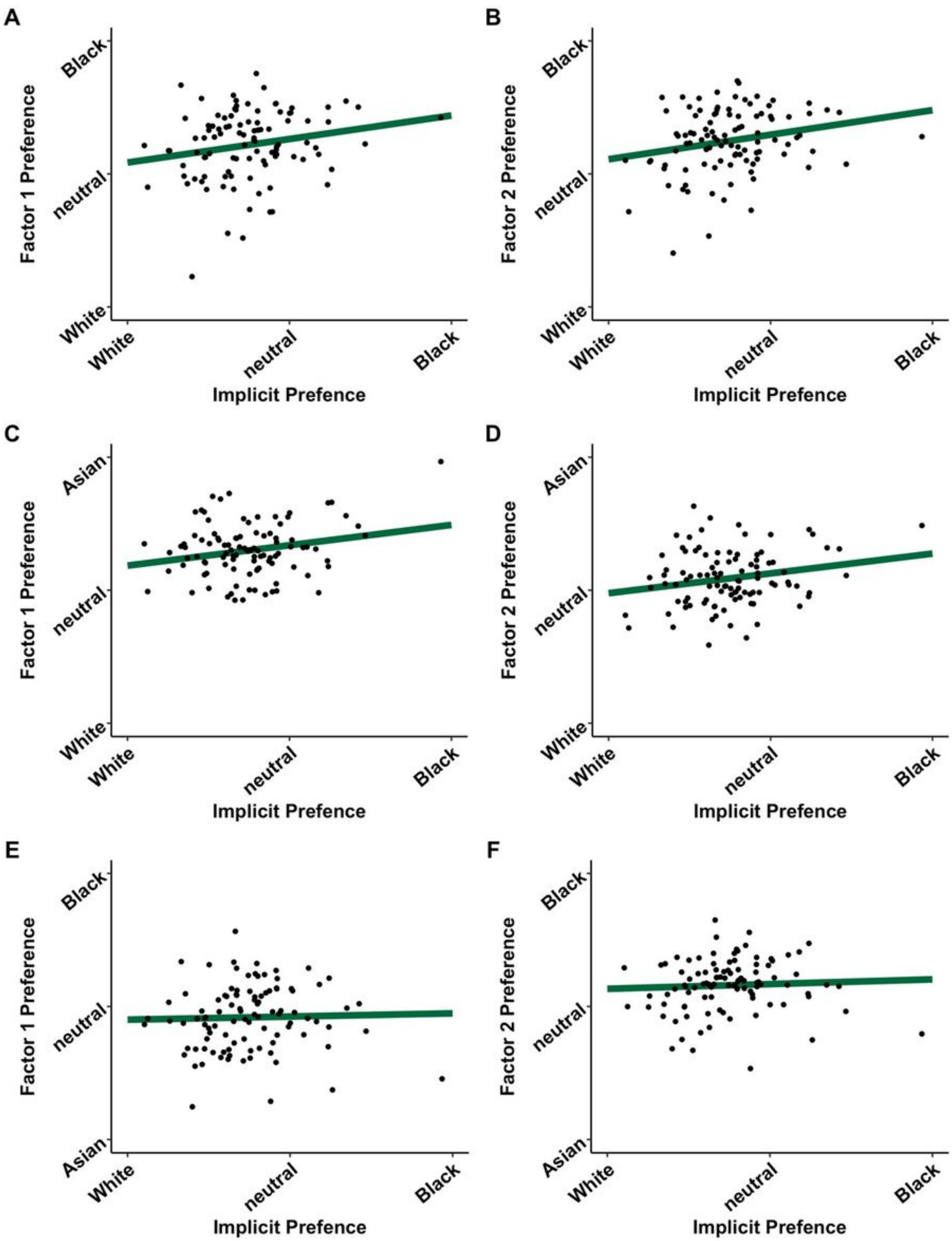

Figure 4

Results from experiment 2. Panels A-F show the relationship of implicit bias measured using the IAT (D, x-axes) to biased trait inferences (y-axes) for Black versus White faces (A & B) Asian versus White faces (C & D) and Asian versus Black faces (E & F). All axes are labelled to indicate which preference positive and negative scores indicate. Each black dot represents one participant, the line represents the best linear fit of implicit to explicit bias scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- exp1data.csv
- exp2data.csv
- supplementaryMaechleretal2022.docx