

**EXPLICIT AND IMPLICIT PERCEPTIONS OF FACIAL TRUSTWORTHINESS:  
WHITE AND BLACK PERSPECTIVES IN AN INTERGROUP CONTEXT**

HANNAH TRAN

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY  
YORK UNIVERSITY  
TORONTO, ONTARIO

SEPTEMBER 2024

© HANNAH TRAN, 2024

## **Abstract**

Previous research demonstrates that our initial impressions of a person's trustworthiness are influenced by the presence of meaningful combinations of facial features perceived as trustworthy (e.g., large round eyes, mouths with upturned corners). Although trustworthiness is crucial to interpersonal interactions, it may be even more important in racial intergroup contexts, wherein misunderstandings and miscommunications can fuel conflict and mistrust between groups. Across four experiments, I examined whether the way that facial cues influence evaluations of trustworthiness differs depending on the race of the target (the person being evaluated) and the race of the perceiver (the person doing the evaluation). Although both White and Black perceivers used facial cues to assess trustworthiness, the impact of target race varied depending on whether the task involved implicit or explicit responding. White participants explicitly rated Black compared to White faces as more trustworthy, but implicitly associated Black faces with untrustworthiness. Black participants also explicitly rated Black faces as more trustworthy, but target race did not impact their implicit trustworthiness judgements. The potential role of social concerns in intergroup trustworthiness evaluations and methodological implications for future research on perceptions of facial trustworthiness are discussed.

*Keywords:* face perception, person perception, intergroup relations, facial trustworthiness

## Acknowledgements

To my supervisor Dr. Kerry Kawakami, thank you for your guidance, direction, and passion for this work, and for all you have done to help me grow as a researcher. It is an unexplainable joy and privilege to be able to learn from you and research alongside you.

To Dr. Cindel White, I am indebted to your generous time and flexibility during the writing of my thesis. To my entire examining committee, including Dr. Joey Cheng and Dr. Winny Shen, thank you for your valuable time and thoughtful comments during my defence.

To Dr. Justin Friesen, thank you for your efforts in recruiting participants and the extensive advice you offered me throughout this work. To Dr. Ronda Lo, Miranda Too, and Dr. Corey Petsnik, thank you for your willingness to share your statistical and R expertise with me.

To Zoe Fagnoli Brown, Courtney Chan, Emily Bissada, Kev Zhou, and later on, Dr. Kunalan Manokara, thank you for making our lab feel like a second home. I could have never imagined when I began graduate school that I would find the level of friendship, support, and collaboration we have all shared with one another. You are all the best of the best.

To the many research assistants in the Social Cognition Lab, thank you for the genuine enthusiasm and earnestness with which you approached this research. This thesis would not have been possible without your tireless hours of running participants and invaluable work.

To my family, near and far, thank you for being there. Thank you to my father for pushing me to spread my wings (your words, not mine); to my grandparents for your immeasurable support; to my brother for keeping things real; to my mother for sharing my interests; to the Yee family for encouraging me like one of your own; to my best friend, Reed, for being an unwavering constant in my life; and finally, to my partner in love and life, Aeydan, for always keeping me grounded. I could not have done any of this without you.

## Table of Contents

Abstract .....	ii
Acknowledgements .....	iii
Table of Contents .....	iv
List of Figures .....	vi
Introduction .....	1
Facial Trustworthiness .....	2
Trustworthiness in a Racial Intergroup Context .....	3
Intergroup Social Concerns .....	7
Explicit and Implicit Measures .....	9
Methods of Current Research .....	10
Current Research .....	12
Experiment 1	
Method .....	14
Results and Discussion .....	16
Experiment 2	
Method .....	18
Results and Discussion .....	19
Experiment 3	
Method .....	23
Results and Discussion .....	24
Experiment 4	
Method .....	26

Results and Discussion .....	27
General Discussion .....	29
Social Concerns as a Potential Explanation of Results .....	31
The Impact of Participant Race.....	34
Methodological Implications .....	38
Conclusion .....	45
References .....	48
Appendices.....	64

## List of Figures

Figure 1: Example of a White Face Depicting a Facial Model of Trustworthiness from Low (-3) to High (+3) Increasing by One Standard Deviation .....	2
Figure 2: Examples of Individual Participants' Mental Representations of Perceived Trustworthiness.....	5
Figure 3: Examples of Stimuli Depicting High and Low Trustworthiness on Black and White Faces .....	10
Figure 4: Example of a Single Trial of the Mouse-Tracking Task .....	11
Figure 5: White Participants' Trustworthiness Ratings for Black and White Targets with High and Low Facial Trustworthiness in Experiment 1 .....	16
Figure 6: Observed Mean AUCs for One White Participant Categorizing White and Black Faces High in Facial Trustworthiness on the Mouse-tracking Task .....	20
Figure 7: White Participants' Areas under the Curve (AUCs) for Black and White Targets with High and Low Facial Trustworthiness in Experiment 2 .....	22
Figure 8: Black Participants' Trustworthiness Ratings for Black and White Targets with High and Low Facial Trustworthiness in Experiment 3 .....	25
Figure 9: Black Participants' Areas under the Curve (AUCs) for Black and White Targets with High and Low Facial Trustworthiness in Experiment 4 .....	28
Figure 10: Examples of Machine-Learning Generated Faces Modelling Perceived Trustworthiness.....	44

## **Explicit and Implicit Perceptions of Facial Trustworthiness:**

### **White and Black Perspectives in an Intergroup Context**

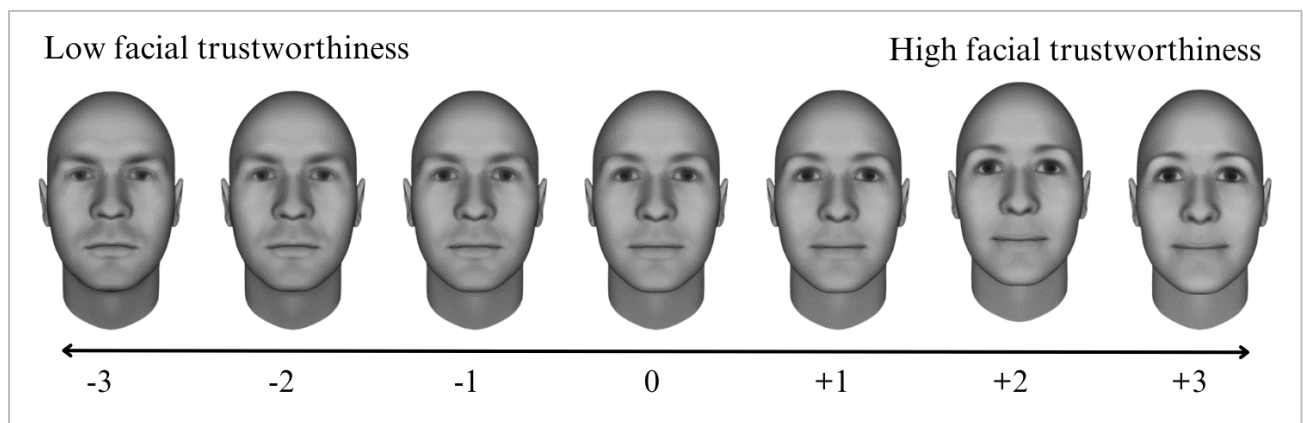
Our rapid, initial impressions of a person's trustworthiness rely heavily on the rich information their face provides (Cloutier et al., 2005), including the morphology of their facial features and the social categories to which they belong. These quick assessments—which can be made within 33 ms of the presentation of faces and stabilize within 100 ms (Todorov et al., 2009)—are vital to understanding interpersonal interactions, given that they can determine how we respond to and treat others (Adams Jr & Kleck, 2005; Hugenberg & Wilson, 2013). Although perceptions of trustworthiness have important downstream consequences (Chang et al., 2010; Friesen & Sinclair, 2010; Fruhen et al., 2015; Porter et al., 2010; Slepian et al., 2017; Stirrat & Perrett, 2010; Sutherland et al., 2020; Thierry & Mondloch, 2021; Wilson & Rule, 2015), it remains unclear whether the way that facial cues influence our evaluations of trustworthiness differs depending on the race of the target (the person being evaluated) and the perceiver (the person doing the evaluation). Traditionally, research on face perception has focused on White targets and White perceivers (Jones et al., 2021), although more recent research has highlighted how target race can have a significant impact when forming impressions (Friesen et al., under review, Kawakami et al., in press; Xie et al., 2021). Investigations of perceiver race, however, have been notably fewer (Kawakami et al., 2017). There is also reason to believe that these processes may differ by conditions that elicit explicit and implicit responding (Bijlstra et al., 2010; Dovidio et al., 2009; Kawakami et al., 2017; Oswald & Adams, 2023). Thus, in the current research, I aimed to extend previous literature by investigating how both target and perceiver race impact perceptions of facial trustworthiness using both explicit and implicit measures.

## Facial Trustworthiness

When forming social impressions, trustworthiness, along with dominance, has been identified as a key dimension of evaluating faces (Oosterhof & Todorov, 2008; Todorov, 2017). With data-driven methods, researchers have been able to build mathematical representations of perceived facial traits by manipulating the shape and texture of artificial, three-dimensional faces (Olivola et al., 2014; Singular Inversions, 2024; Todorov et al., 2008). Based on facial trait judgements, researchers have constructed facial models that demonstrate how changes in facial features increase or decrease participant perceptions of particular traits (Hutchings et al., 2024). One well-documented facial model depicting White faces has demonstrated that meaningful combinations of facial features can impact perceived trustworthiness (Todorov et al., 2011, 2013), see Figure 1.

**Figure 1**

*Example of a White Face Depicting a Facial Model of Trustworthiness from Low (-3) to High (+3) Increasing by One Standard Deviation*





Specifically, this model indicates that feature combinations related to high trustworthiness include larger, rounder eyes, a larger mouth with upturned corners, pronounced cheekbones, high inner eyebrows, a shallow nose sellion, and a wide, shorter chin, while opposing features are related to low trustworthiness (Todorov & Oosterhof, 2011). In general, faces perceived as trustworthy tend to contain features that are more feminine, baby-faced, and reflect positive emotions, while faces perceived as untrustworthy tend to be more masculine, older, and reflect negative emotions (Hutchings et al., 2024; Olivola et al., 2014).

Although perceptions of facial trustworthiness are commonly used to make inferences regarding targets' intentions as harmful or harmless, these superficial features are not a reliable indicator of their inner traits (Olivola et al., 2014; Todorov et al., 2015). Nonetheless, these judgements are important, given that they lead to a variety of downstream consequences. For example, having a face that is perceived as trustworthy has been associated with a higher likelihood of being given a financial loan (Duarte et al., 2012), higher pay in the workplace at the managerial level (Fruhen et al., 2015), and a higher likelihood of being trusted in experimental strategic economic games, even in the face of relevant information about past trustworthiness-related behaviours (Chang et al., 2010). Similarly, having a face perceived as less trustworthy has been associated with serious consequences in the legal system, such as participants requiring less evidence to convict such targets of a crime (Porter et al., 2010).

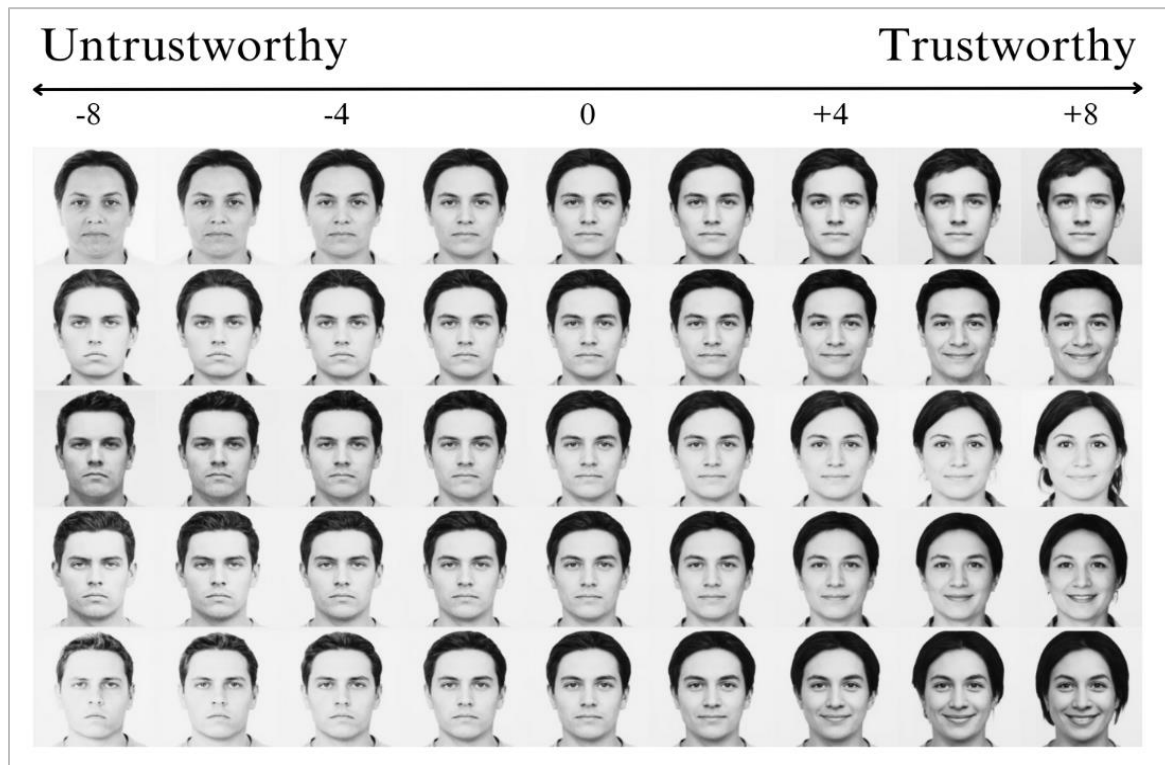
### **Trustworthiness in a Racial Intergroup Context**

Previous research has established that at the universal level, perceiver consensus is high when assessing trustworthiness (Rule et al., 2012; Todorov & Oosterhof, 2013), with one study suggesting that the trustworthiness-dominance model generalizes across most regions of the world (Jones et al., 2021). More recent research, however, demonstrates that perceptions of

trustworthiness can also be idiosyncratic. For example, using a data-driven visualization method, researchers had participants rate a series of machine-learning-generated faces as trustworthy, untrustworthy, or neither (Albohn et al., 2022). Using these ratings and the same neutral starting point for each participant, researchers were able to create a facial trustworthiness model for each individual that corresponded to their mental representations of faces ranging from low (-8) to high (+8) in trustworthiness, see Figure 2. Although there was some consensus about what a highly trustworthy face looked like (e.g., generally happier and younger), participants' mental representations of highly trustworthy and untrustworthy faces appeared visually distinct from one another on several dimensions including sex and facial morphology.

**Figure 2**

*Examples of Individual Participants' Mental Representations of Perceived Trustworthiness*



*Note.* Each row represents one participant, while each column represents a single directional vector from -8 to +8. Figure adapted from “A data-driven, hyper-realistic method for visualizing individual mental representations of faces” by Albohn, D. N., Uddenberg, S., & Todorov, A., 2022, *Frontiers in Psychology*, p. 10.

While previous research has examined both universal and individual perceptions of facial trustworthiness, the impact of group-level differences in this process has yet to be thoroughly investigated (Kawakami et al., 2017). Although the ability to process facial cues and assess trustworthiness is vital to interpersonal interactions and outcomes, it may be even more important in a racial intergroup context, where misunderstandings and miscommunications can

fuel conflict, suspicion, and mistrust between groups (Brewer, 1999; Dovidio et al., 2002; Lloyd et al., 2017; Tropp, 2008).

When considering target race alone, researchers have provided strong evidence for its influence on participants' trait perceptions (Hugenberg & Bodenhausen, 2003; Jaeger & Jones, 2021; Karmali & Kawakami, 2023). This difference, however, often goes unobserved in typical research paradigms wherein target race is held constant in favour of other variables of interest (Jaeger & Jones, 2021). For example, the literature on facial trustworthiness perceptions has largely relied on White faces (e.g., Collova et al., 2019; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Swe et al., 2020; Todorov et al., 2009, 2013; Vernon et al., 2014; Willis & Todorov, 2006). Existent facial trustworthiness models, therefore, have been created and validated primarily on White faces (Todorov et al., 2013). Notably, whereas prior research indicates that participants can distinguish between subtle variations of high and low facial trustworthiness on White targets, it is currently unclear how this process generalizes to Black target stimuli (Todorov et al., 2011).

Perceptions of target race may also differ depending on the race of the perceiver. That is, perceivers from different races may have unique lived experiences relating to their race that impact their perceptual processes and evaluations (Bijlstra et al., 2010; Oswald & Adams, 2023; Roberts et al., 2020). Given that investigations of perceiver race in studies of facial trustworthiness have been relatively rare (Kawakami et al., 2017, in press), it remains unclear how perceivers' racial group membership might affect perceptions of trustworthiness (Todorov et al., 2009, 2013). Notably, although facial models of trustworthiness have been validated with homogenous, primarily White samples (Todorov et al., 2013), these models' efficacy with other-race samples has yet to be explored.

Critically, the race of the target and the race of the perceiver may interact when forming trustworthiness perceptions, given that perceivers' evaluations in interracial interactions may differ depending on whether they belong to a majority (i.e., White) or minority (e.g., Black) racial group (Cohen & Steele, 2002; Holoien et al., 2015; Vorauer, 2005). For example, research demonstrates that Black people approach interracial interactions with White people with mistrust, guardedness, and anxiety (Dovidio et al., 2002; Shelton, 2000; Swim et al., 1998). Additionally, perceivers' racial beliefs and experiences impact the extent to which they perceive faces as trustworthy. For example, White people who hold racial stereotypes such as believing that White people as a whole are highly trustworthy are more likely to evaluate a White face as trustworthy relative to other-race faces (Xie et al., 2021), and White people holding anti-Black prejudice and who report low levels of intergroup contact have a higher tendency to judge prototypically White features as trustworthy and prototypically Black features as untrustworthy (Hutchings et al., 2024). Given the importance of both target and perceiver race in such processes, the current research sought to extend the literature by examining how the races (i.e., White or Black) of both the target and the perceiver impact perceptions of facial trustworthiness.

### **Intergroup Social Concerns**

Given that perceivers' lived experiences and socialization can affect perceptual processes, it is necessary to consider how distinct social concerns related to these experiences can impact intergroup relations. Research indicates that interracial interactions are perceived as stressors in different ways for different groups (Richeson & Shelton, 2007). Specifically, majority group members (e.g., White people) often experience stress due to concerns that they are behaving in prejudiced ways, while minority group members (e.g., Black people) experience

stress due to concerns that they are the target of racial prejudice and are behaving in ways that confirm prejudiced expectations.

Not surprisingly, White and Black people engage in different strategies to cope with these different concerns. Given concerns about appearing prejudiced and strong societal norms discouraging expressions of racial bias (Devine et al., 2002; Kawakami et al., 2017; Shelton, 2000), White people may attempt to not behave in ways that suggest that they are treating people from other racial groups differently. For example, despite the persistence of racism in contemporary society (Roberts et al., 2020; Roberts & Rizzo, 2021), self-reported responses from White participants often do not indicate prejudice (Dovidio et al., 2002). Furthermore, previous research demonstrates that White people's social concerns about appearing racist can drive them to adopt colour-blind strategies, which promote the idea that not acknowledging race makes one appear unprejudiced (Karmali et al., 2019). Thus, when it comes to reporting trustworthiness perceptions, White perceivers may adopt similar strategies and explicitly report perceiving no differences in trustworthiness between Black and White target faces.

However, Black people's experiences as the targets of racism may socialize them to engage in strategies that reduce the impact of racial bias. Whereas White people may be socialized to ignore racial differences between racial groups, Black people are socialized to emphasize this divide (Shelton, 2000). Although Black people may use strategies in interracial interactions such as behaving in an overly positive manner to facilitate amicable interactions (Richeson & Shelton, 2007), they may also cope with being the target of prejudice by seeking social support from other Black people and extending support in return (Barr & Neville, 2014; Johnson & Carter, 2019; Tang et al., 2016). Furthermore, Black cultural norms traditionally emphasize the importance of holding positive attitudes about other Black people (Helms, 1990;

Sellers et al., 1998). For example, the Multidimensional Model of Racial Identity highlights Black people's private regard (i.e., the extent to which one feels positively) for other Black people as vital to Black racial identity and behaviours, particularly in contexts where race is made salient (Sellers et al., 1998). Thus, in the intergroup context of perceiving facial trustworthiness on Black and White targets, when it comes to reporting trustworthiness perceptions, Black perceivers may explicitly report perceiving higher trustworthiness on Black than White target faces.

### **Explicit and Implicit Measures**

In typical studies examining facial trustworthiness perceptions, participants are asked to evaluate faces using Likert scales (Olivola, 2014; Todorov et al., 2013), which capture their explicit responses (Gawronski & Bodenhausen, 2006). Because of participants' abilities to adjust and monitor their responses on such measures, their evaluations may be influenced by social concerns (Kawakami et al., 2017). In contrast, few studies on facial trustworthiness have evaluated responses on implicit measures, which reflect more automatic and spontaneous associations with social categories and may be less influenced by social concerns (Dovidio et al., 2009; Gawronski & Bodenhausen, 2006).

Although previous literature suggests that White people respond positively toward Black people on explicit measures (Devine et al., 2002), their responses on more negative implicit measures have been more negative (Brown-Iannuzzi et al., 2019; Hugenberg & Bodenhausen, 2003; Kubota & Ito, 2014; March et al., 2021). It is less clear, however, how and whether Black people will differ in their perceptions of trustworthiness on explicit and implicit measures (Richeson & Shelton, 2007; Sellers et al., 1998). Specifically, because racial socialization may encourage the importance of supporting other Black people and a shared history of

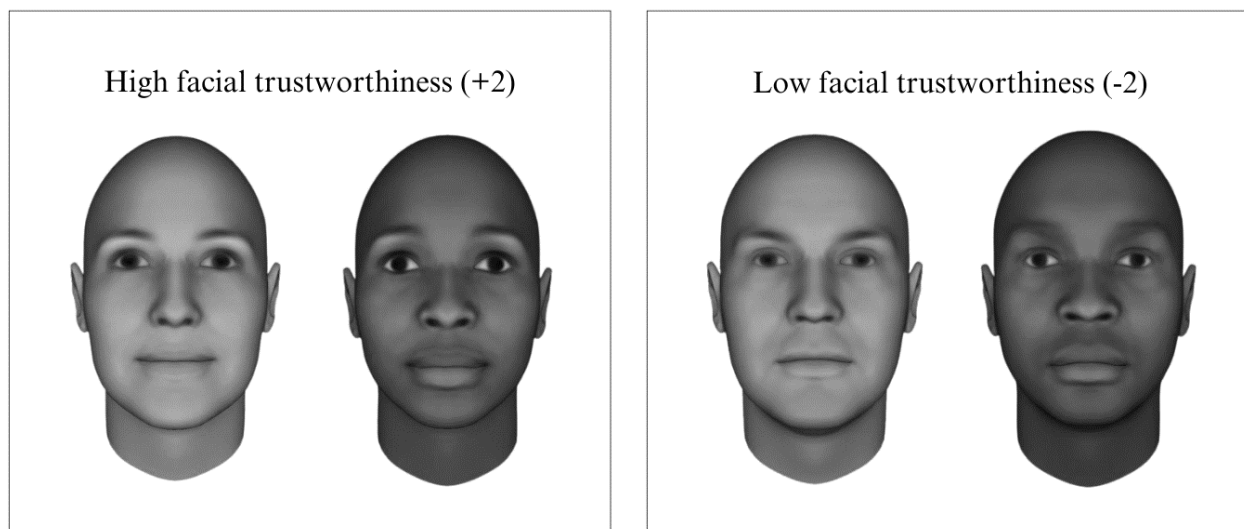
discrimination may facilitate ingroup solidarity (Craig & Richeson, 2016; Roberts et al., 2020; Rotella et al., 2013), Black people may perceive other Black people more positively and view them positively than outgroups (Ho et al., 2017), at least on explicit measures. It is unclear, however, how Black people will respond to implicit measures. Do they also respond more positively when they are less able to monitor and control their responses? In short, although the different social concerns of White and Black perceivers may affect their explicit evaluations, it is important to further examine how their responses may differ on implicit evaluations.

### Methods of Current Research

In the current research, I included both an explicit and implicit measure to examine perceived facial trustworthiness on faces containing facial cues that were perceptually distinct to participants as high (2 standard deviations above the mean) and low (2 standard deviations below the mean) in facial trustworthiness, see Figure 3 (Todorov et al., 2011, 2013).

### Figure 3

*Examples of Stimuli Depicting High and Low Trustworthiness on Black and White Faces*

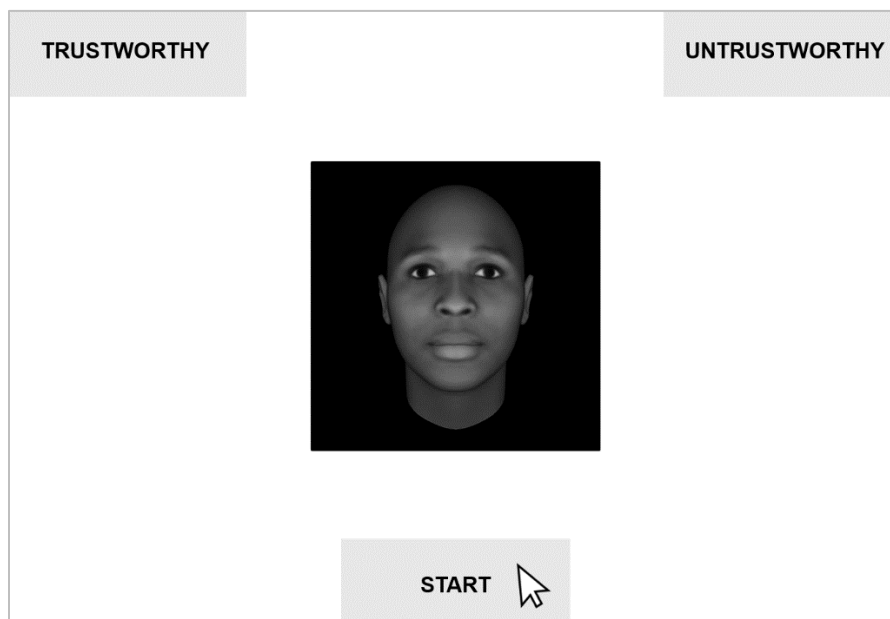




To examine participants' explicit responses, I adopted a rating task used in previous research on trustworthiness (Todorov et al., 2013). Specifically, participants were presented with a set of faces in succession and asked to rate the trustworthiness of each face on a Likert scale. To examine implicit responses, I used a mouse-tracking task paradigm. This well-validated measure uses deviations in participants' hand trajectories to indirectly measure how category-related content is activated during person construal (Freeman & Ambady, 2009; Kawakami et al., 2020). Specifically, I examined the extent to which participants associated Black and White targets with trustworthiness. Importantly, because participants are unaware that their mouse movements are being monitored, responses to this paradigm may be less impacted by social concerns. In a typical two-choice mouse-tracking task, participants use their mouse to categorize target stimuli on a computer screen into one of two response alternatives at the top left and right of their screen (Freeman & Ambady, 2009), see Figure 4.

#### **Figure 4**

*Example of a Single Trial of the Mouse-Tracking Task*



Participants in mouse-tracking tasks typically choose the intended response (e.g., categorizing a face high in trustworthiness as trustworthy). However, when the path of their mouse's trajectory is less direct and deviates toward the unchosen response, it indicates an association with the unchosen category (Freeman et al., 2008). In the current research, if participants' mouse trajectories deviated more toward "untrustworthy" before ultimately choosing "trustworthy" when presented with a Black versus White face high in facial trustworthiness, it would indicate a stronger association between untrustworthiness and Black targets.

### **Current Research**

Across four experiments, the current research examined how perceived facial trustworthiness on White and Black targets is impacted by the extent to which facial features depict high or low trustworthiness and whether perceivers are White or Black. In Experiment 1, the focus was on a task assessing explicit responses with trait ratings and White participants. In accordance with previous research on facial trustworthiness (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2013; Wilson & Rule, 2015), I predicted a main effect of facial trustworthiness in which participants would rate faces high in facial trustworthiness as more trustworthy than faces low in facial trustworthiness. In accordance with White people's social concerns of not appearing racist (Devine et al., 2002; Karmali et al., 2019; Kawakami et al., 2017), I did not expect a main effect or interaction with target race on this measure.

In Experiment 2, the focus was on a task assessing implicit responses with a mouse-tracking task and White participants. Given that implicit measures may be less impacted by social motivations and often reveal racial biases (Dovidio et al., 2009; Kawakami et al., 2017), I expected an interaction effect between facial trustworthiness and target race. Specifically, I

predicted that White participants would be less direct at categorizing Black compared to White targets that were high in facial trustworthiness. I also expected participants to be less direct at categorizing White compared to Black faces that were low in facial trustworthiness.

In Experiment 3, the focus was on a task assessing explicit responses with trait ratings and Black participants. In accordance with previous research on facial trustworthiness (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2013; Wilson & Rule, 2015), I predicted a main effect of facial trustworthiness in which participants would rate faces high in facial trustworthiness as more trustworthy than faces low in facial trustworthiness. In accordance with Black people's social concerns related to positively evaluating other Black people (Helms, 1990; Sellers et al., 1998), I also predicted a main effect of target race. Specifically, I predicted that Black participants would rate Black compared to White faces as more trustworthy. I did not expect an interaction effect.

In Experiment 4, the focus was on a task assessing implicit responses with a mouse-tracking task and Black participants. Because it is unclear whether Black participants' social concerns about evaluating other Black people would positively impact their implicit responses, I had two potential predictions. First, if Black participants' support of other Black people was based on others' beliefs, then I would predict a main effect of facial trustworthiness. Specifically, in accordance with previous research on facial trustworthiness (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2013; Wilson & Rule, 2015), Black participants would be less direct at categorizing faces low compared to high in facial trustworthiness, and I would not expect a main effect or interaction with target race. Alternatively, if Black people have learned to associate trustworthiness with other Black compared to White people (Craig & Richeson, 2016; Ho et al., 2013; Rotella et al., 2013), I would predict an interaction effect between facial

trustworthiness and race. Specifically, I would expect that while Black participants would be less direct at categorizing White compared to Black faces that were high in facial trustworthiness, they would be less direct at categorizing Black compared to White faces that were low in facial trustworthiness.

## **Experiment 1**

### **Method**

#### ***Participants and Design***

Experiment 1 focused on an explicit measure to examine whether White participants' trustworthiness ratings were impacted by the extent to which facial features depicted high and low trustworthiness and whether targets were White or Black. Because prior research has relied primarily on White perceivers and White targets (Todorov et al., 2013), the current experiment extended this literature by examining how target race affected these perceptions.

In Experiment 1, I used a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) within-subjects design to assess ratings of trustworthiness. Using the typical effect size in social and personality psychology ( $r = .21$ ; Funder et al., 2014), an a priori power analysis using G\*Power 3.1 (Faul et al., 2007) suggested that a minimum sample of 86 was needed to detect an interaction between target race and facial trustworthiness assuming a correlation among repeated measures of  $r = .50$  (power = .80,  $\alpha = .05$ ). To ensure adequate power, I aimed to oversample by approximately 15 percent for a final sample size of 100 participants after exclusions. In total, 118 participants completed Experiment 1, as I examined demographic-related exclusion criteria as the data came in and continued to recruit participants to replace those removed through a priori exclusion criteria. Given that the current research examined trustworthiness in a specific intergroup cultural context (Lo & Sasaki, 2023), the data

from participants who had lived in Canada for less than three years were excluded from the analysis ( $n = 2$ ). Because this experiment required participants to pay attention and react quickly, the data from participants who stated that they did not pay attention ( $n = 1$ ) or who reported that they had an attentional disorder (e.g., ADHD;  $n = 15$ ) were also removed from analysis. The final sample included 100 White participants (82 female, 18 male,  $M_{\text{age}} = 18.73$ ,  $SD_{\text{age}} = 4.36$ ).

### ***Procedure***

Upon arriving at the laboratory, participants were informed that the study investigated first impressions. They were instructed to rate a series of faces on trustworthiness and to not take too long in evaluating each face. To eliminate potential stimuli-level differences, participants were randomly assigned to view one of two counterbalanced sets of 48 computer-generated faces two standard deviations above and below the mean in facial trustworthiness (Singular Inversions, 2024; Todorov et al., 2011, 2013; see Appendix A). Based on the available stimuli, all faces were male. Half of the faces were White, and half the faces were Black. Although each set contained all 48 faces, only one version (high or low in facial trustworthiness) of each face was included in each set so that participants did not see the same face twice. All faces were presented in a different random order for each participant.

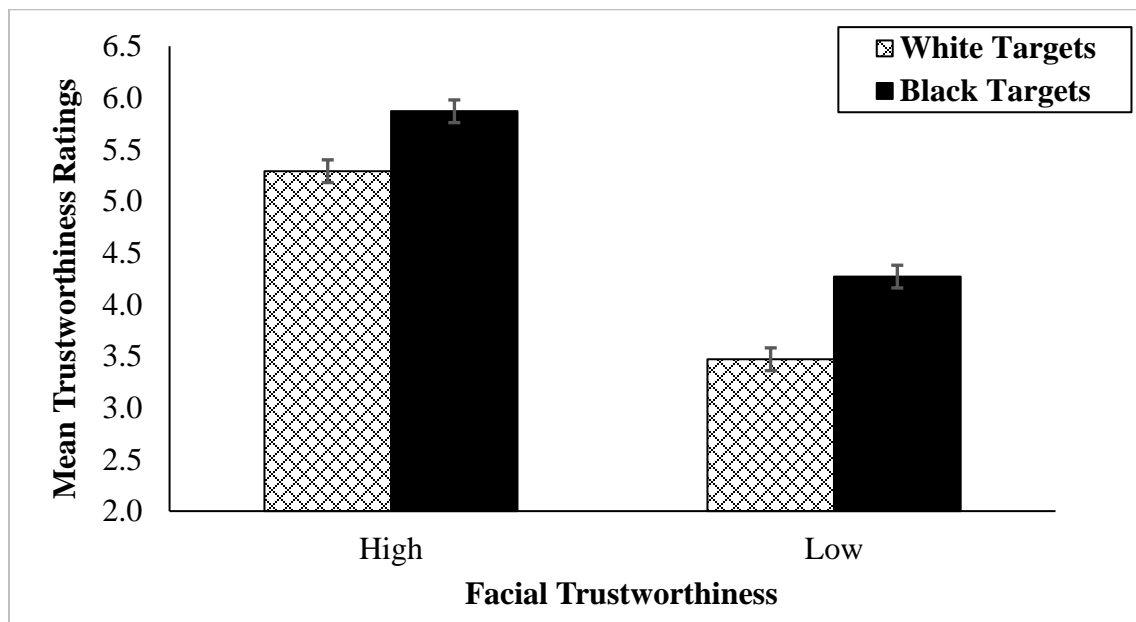
On each trial, a face appeared for 3000 ms, after which participants rated its trustworthiness from 1 (not at all trustworthy) to 9 (very trustworthy). Participants then completed a set of demographic questions (see Appendix B) as well as a set of questions regarding their experience during the experiment (e.g., what they believed the researchers were predicting, whether they were distracted during the experiment; see Appendix C).

## Results and Discussion

A 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) repeated-measures analysis of variance (ANOVA) was performed on participants' mean trustworthiness ratings. The main effect of facial trustworthiness was significant,  $F(1, 99) = 407.49, p < .001, \eta_p^2 = 0.80$ , 95% CI [0.74, 0.85], with faces high ( $M = 5.58, SD = 1.16$ ) compared to low ( $M = 3.87, SD = 1.25$ ) in facial trustworthiness rated as more trustworthy. The main effect of target race was also significant,  $F(1, 99) = 70.39, p < .001, \eta_p^2 = 0.42$ , 95% CI [0.27, 0.53], with Black ( $M = 5.07, SD = 1.42$ ) compared to White ( $M = 4.38, SD = 1.46$ ) faces rated as more trustworthy. Notably, these effects were qualified by a two-way interaction,  $F(1, 99) = 5.17, p = .025, \eta_p^2 = 0.05$ , see Figure 5.

**Figure 5**

*White Participants' Trustworthiness Ratings for Black and White Targets with High and Low Facial Trustworthiness in Experiment 1*



Pairwise comparisons indicated that when targets were high in facial trustworthiness, participants rated Black ( $M = 5.87$ ,  $SD = 1.12$ ) compared to White ( $M = 5.29$ ,  $SD = 1.22$ ) faces as significantly more trustworthy,  $t(99) = 5.94$ ,  $p < .001$ ,  $d = 5.94$ , 95% CI [0.42 to 11.68]. When targets were low in facial trustworthiness, participants also rated Black ( $M = 4.27$ ,  $SD = 1.23$ ) compared to White ( $M = 3.47$ ,  $SD = 1.15$ ) faces as more trustworthy,  $t(99) = 8.53$ ,  $p < .001$ ,  $d = 8.53$ , 95% CI [0.98 to 16.58]. A paired-samples t-test on the difference scores between trustworthiness ratings for Black and White faces indicated that the effect was larger when targets were low ( $M = 0.80$ ,  $SD = 0.94$ ) compared to high ( $M = 0.58$ ,  $SD = 0.97$ ) in facial trustworthiness,  $t(99) = -2.27$ ,  $p = .025$ ,  $d = -0.23$ , 95% CI [-0.42, -0.03].

Consistent with prior literature on perceptions of facial trustworthiness, the results of Experiment 1 demonstrated that participants used facial cues to determine perceived trustworthiness, with participants rating faces high compared to low in facial trustworthiness as more trustworthy (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2011, 2013). Experiment 1, however, extended this research by demonstrating that target race impacts this process.

In accordance with findings suggesting that White people may be influenced by concerns related to appearing racist when responding to explicit measures (Dovidio et al., 2009; Kawakami et al., 2017; Shelton, 2000), White participants rated Black compared to White faces as more trustworthy. Notably, when presented with cues that Black targets were not trustworthy (i.e., low facial trustworthiness), this effect was even larger. Despite the suggestion that facial trustworthiness is a universal process (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2011, 2013), these initial findings indicate that perceiver and target race might also impact these perceptions. Although I initially predicted that there would be no difference in

trustworthiness ratings based on target race, it is notable that White perceivers were even more positive toward Black than White targets, especially when negative evaluations were implicated (i.e., low trustworthiness cues). In Experiment 2, I extended these findings by examining how target race impacts White participants' perceptions of trustworthiness using an implicit measure.

## **Experiment 2**

### **Method**

#### ***Participants and Design***

Experiment 2 focused on an implicit measure to examine whether White participants' trustworthiness responses were impacted by the extent to which facial features depicted high or low trustworthiness and whether targets were White or Black. Given that prior research indicates that explicit measures may be affected by participants' social motivations and concerns (Devine et al., 2002; Karmali et al., 2019), in Experiment 2, I utilized a mouse-tracking paradigm to reduce participants' abilities to monitor and modify their responses.

In Experiment 2, I used a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) within-subjects design to assess the categorization of trustworthiness. In accordance with Experiment 1, a sample of 86 was needed to detect an interaction between target race and facial trustworthiness. To ensure adequate power, 132 White participants were recruited from the undergraduate participant pool in exchange for course credit using the same sampling and exclusion methods as Experiment 1. The data from participants who had lived in Canada for less than three years ( $n = 3$ ), who did not read the instructions or complete the study ( $n = 3$ ), who reported that they had an attentional disorder (e.g., ADHD;  $n = 25$ ), and who engaged in invariant responding ( $n = 6$ ) were excluded. The final sample included 95 White (76 female, 19 male,  $M_{\text{age}} = 19.21$ ,  $SD_{\text{age}} = 2.58$ ) participants.



## ***Procedure***

Upon arriving at the laboratory, participants were informed that the study investigated how faces were perceived based on trustworthiness. They were instructed to respond to a series of faces and asked to categorize whether faces were trustworthy or untrustworthy by using a computer mouse. Participants were randomly assigned to one of two sets of 48 faces described in Experiment 1. On each trial, a “start” button appeared at the bottom centre of the screen. For half of the participants, the word “trustworthy” appeared at the top left and the word “untrustworthy” at the top right of the screen, while for the other half of the participants, the placement of the labels was reversed. To begin a trial, participants used their mouse to click on “start,” after which a target face appeared, and participants used their mouse to select one of the two response alternatives. The target face disappeared either after a maximum of 2000 ms or as soon as the participant responded. If participants did not start moving their mouse within 400 ms of clicking on “start,” a message appeared on the screen instructing them to begin moving as quickly as possible in future trials. Before completing the experimental trials, participants completed four practice trials with an experimenter present to answer any questions and provide directive. After completing the mouse tracking task, participants answered the same demographic and study experience questions used in Experiment 1.

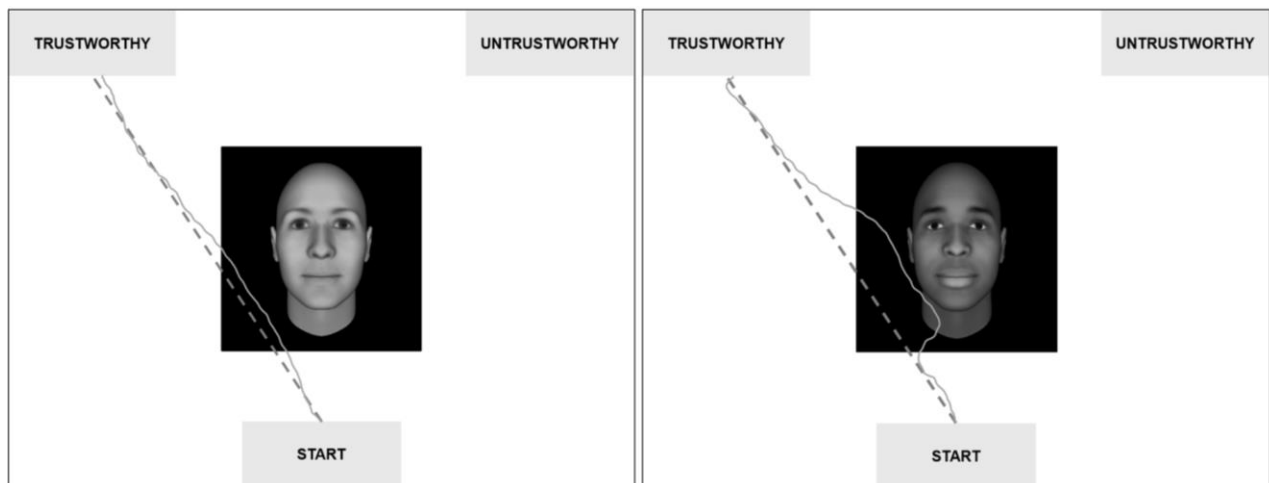
## **Results and Discussion**

To examine the impact of facial trustworthiness cues and target race cues on the categorization of faces as either trustworthy or untrustworthy, I calculated the mean area under the curve (AUC) for each participant. The AUC is the geometric area between the idealized response trajectory (i.e., a straight line between “start” and the correct response option) and the participant’s actual trajectory from “start” to response (Freeman & Ambady, 2010). A large

AUC (i.e., a less direct trajectory) indicates a stronger association with the unselected response. For example, if a participant had a larger AUC for faces high in facial trustworthiness when they were Black compared to White, it would indicate that they held a stronger association between Black faces and untrustworthiness; see Figure 6 for an observed individual-level AUC with this pattern.

### Figure 6

*Observed Mean AUCs for One White Participant Categorizing White and Black Faces High in Facial Trustworthiness on the Mouse-tracking Task*



*Note.* The dotted line represents the idealized response trajectory (i.e., a direct line from “start” to “trustworthy”), while the solid line represents a participant’s actual mean trajectory across all trials for White (left) and Black (right) targets.

In accordance with previous research on mouse-tracking, only “correct” responses were analyzed (Freeman et al., 2011). A response was considered correct when it was congruent with facial trustworthiness cues (e.g., a face high in facial trustworthiness was categorized as

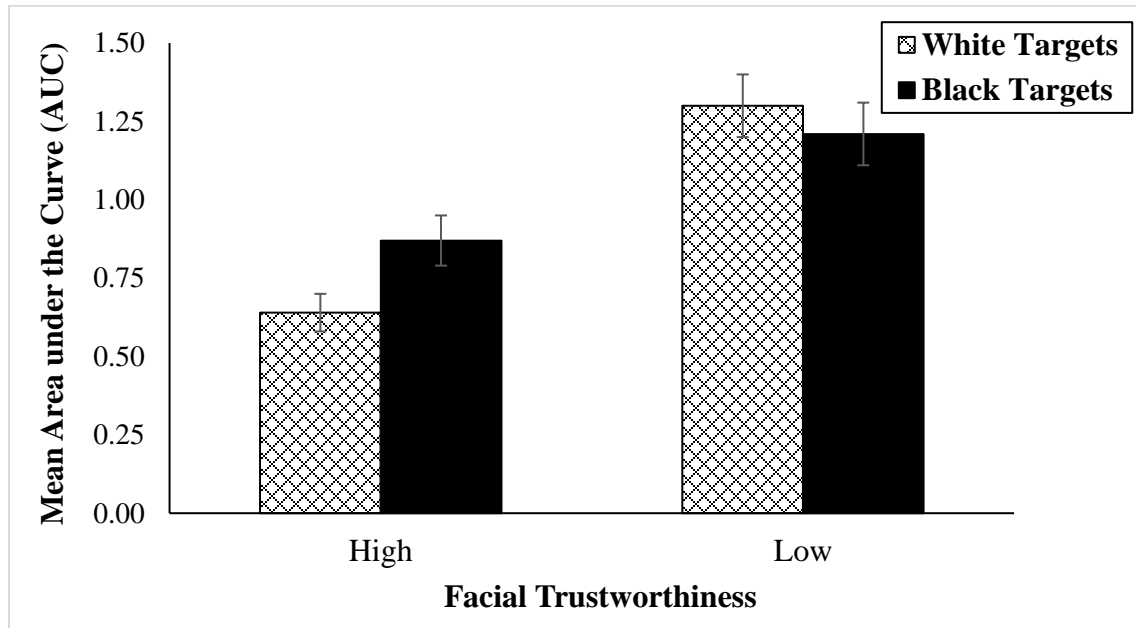
trustworthy or a face low in facial trustworthiness was categorized as untrustworthy).

Participants selected the correct response in 74% of all trials. In accordance with previous research (Freeman et al., 2009), trials were also removed if participants had an initial reaction time of less than 400 ms or a final reaction time of over 2000 ms.

A 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) repeated-measures analysis of variance (ANOVA) was performed on participants' AUCs. The main effect of facial trustworthiness was significant,  $F(1, 94) = 16.95, p < .001, \eta_p^2 = 0.16, 95\% \text{ CI } [0.05, 0.30]$ , with larger AUCs for faces low ( $M = 1.26, SD = 0.69$ ) compared to high ( $M = 0.75, SD = 0.98$ ) in facial trustworthiness. The main effect of target race was not statistically significant,  $F(1, 94) = 3.89, p = .052, \eta_p^2 = 0.04, 95\% \text{ CI } [0.00, 0.15]$ , with no difference in AUCs for Black faces ( $M = 1.03, SD = 0.90$ ) compared to White faces ( $M = 0.97, SD = 0.87$ ). Importantly, the two-way interaction was significant,  $F(1, 94) = 5.19, p = .025, \eta_p^2 = 0.06$ , see Figure 7.

**Figure 7**

*White Participants' Areas under the Curve (AUCs) for Black and White Targets with High and Low Facial Trustworthiness in Experiment 2*



Pairwise comparisons indicated that, for faces high in facial trustworthiness, AUCs were larger for Black ( $M = 0.87$ ,  $SD = 0.76$ ) compared to White ( $M = 0.64$ ,  $SD = 0.61$ ) faces,  $t(87) = 3.65$ ,  $p = .002$ ,  $d = 3.65$ , 95% CI [-0.19, 7.45]. However, for faces low in facial trustworthiness, there was no statistically significant difference in AUCs between Black ( $M = 1.21$ ,  $SD = 1.00$ ) and White ( $M = 1.3$ ,  $SD = .97$ ) faces,  $t(87) = -0.57$ ,  $p = .941$ ,  $d = -0.57$ , 95% CI [-2.54, 1.52].

In Experiment 2, I used an implicit measure that was less vulnerable to social motivations and self-presentation concerns that may impact explicit measures (Kawakami et al., 2017). In line with my prediction, White participants were less direct at categorizing faces high in facial trustworthiness when targets were Black compared to White. These results indicated that even when facial cues on Black faces depicted high trustworthiness, responses deviated toward

untrustworthiness. Although I also predicted that when faces were low in facial trustworthiness, White participants would be less direct at categorizing White compared to Black targets, mouse trajectories did not differ with these images.

In accordance with previous research indicating that White people hold negative implicit associations with Black people (Brown-Iannuzzi et al., 2019; Hugenberg & Bodenhausen, 2003; Kubota & Ito, 2014; March et al., 2021), the present findings suggest that White participants' initial tendencies are to categorize Black compared to White faces as untrustworthy. Although participants corrected these tendencies when responding to Black faces high in facial trustworthiness by eventually selecting the correct response (i.e., ultimately categorizing Black faces high in facial trustworthiness as trustworthy), it is notable that they did not differ in the directness of their mouse trajectories when categorizing Black and White faces low in facial trustworthiness. In Experiments 3 and 4, I extended the findings from Experiments 1 and 2 by examining how perceiver race impacts these processes with Black participants.

### **Experiment 3**

#### **Method**

Experiment 3 focused on an explicit measure to examine whether Black participants' trustworthiness ratings were impacted by the extent to which facial features depicted high or low trustworthiness and whether targets were White or Black. This research extends prior research on facial trustworthiness perceptions that has largely ignored perceiver race in the investigation of these perceptual processes (Kawakami et al., 2017, in press).

In Experiment 3, I used a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) within-subjects design to assess Black participants' explicit ratings of trustworthiness. In accordance with previous experiments, a sample of 86 was needed to detect

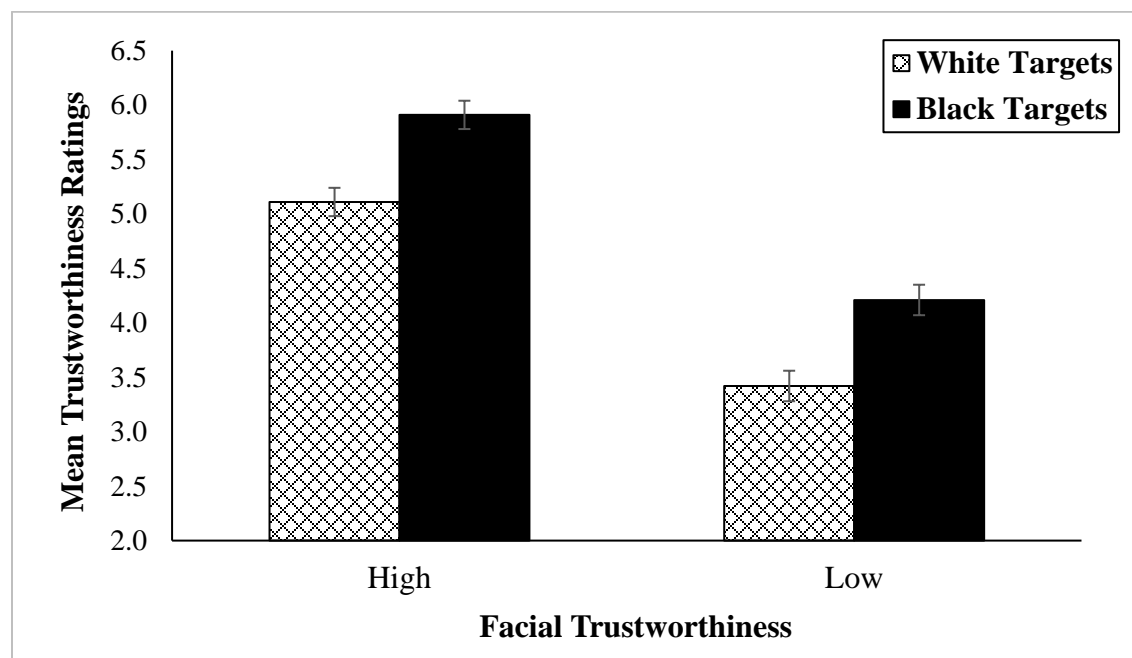
an interaction between target race and facial trustworthiness. To ensure adequate power, 108 Black participants were recruited from the undergraduate participant pool in exchange for course credit using the same sampling and exclusion methods as previous experiments. The data from participants who had lived in Canada for less than three years ( $n = 9$ ) and who did not read the instructions or complete the study ( $n = 4$ ) were excluded. The final sample included 95 Black participants (68 female, 26 male, 1 agender,  $M_{\text{age}} = 19.72$ ,  $SD_{\text{age}} = 4.36$ ). The procedure was identical to that of Experiment 1.

## Results and Discussion

A 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) repeated-measures analysis of variance (ANOVA) was performed on participants' mean trustworthiness ratings. The main effect of facial trustworthiness was significant,  $F(1, 94) = 232.61$ ,  $p < .001$ ,  $\eta_p^2 = 0.71$ , 95% CI [0.62, 0.78], with faces high ( $M = 5.51$ ,  $SD = 1.34$ ) compared to low ( $M = 3.81$ ,  $SD = 1.20$ ) in facial trustworthiness rated as more trustworthy. The main effect of target race was also significant,  $F(1, 94) = 84.26$ ,  $p < .001$ ,  $\eta_p^2 = 0.47$ , 95% CI [0.33, 0.59], with Black ( $M = 5.06$ ,  $SD = 1.51$ ) compared to White faces ( $M = 4.27$ ,  $SD = 1.45$ ) rated as more trustworthy. Notably, the two-way interaction was not significant,  $F(1, 94) = 0.03$ ,  $p = .852$ ,  $\eta_p^2 < .01$ , see Figure 8.

**Figure 8**

*Black Participants' Trustworthiness Ratings for Black and White Targets with High and Low Facial Trustworthiness in Experiment 3*



The results of Experiment 3 demonstrated that, consistent with prior literature with White perceivers, Black participants used facial cues related to trustworthiness in their perceptions, rating targets high compared to low in facial trustworthiness as more trustworthy (Rule et al., 2012; Todorov & Oosterhof, 2011; Todorov et al., 2011, 2013). Furthermore, in accordance with literature suggesting that Black social norms encourage positive evaluations of Black people (Helms, 1990; Sellers et al., 1998), I found that Black participants rated Black compared to White faces as more trustworthy. Notably, the level of facial trustworthiness cues did not qualify these findings. Regardless of whether faces were high or low in facial trustworthiness, Black participants rated Black compared to White targets as more trustworthy. In Experiment 4, I

extended these findings by examining how target race impacts trustworthiness perceptions for Black participants on an implicit measure.

### **Experiment 4**

#### **Method**

Experiment 4 focused on an implicit measure to examine whether Black participants' trustworthiness responses were impacted by the extent to which facial features depicted high or low trustworthiness and whether targets were White or Black. By using a mouse-tracking paradigm, this experiment extended the literature on facial trustworthiness perceptions both with a task assessing implicit attitudes and by examining how perceiver race impacts Black participants' perceptual processes.

In Experiment 4, I used a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) within-subjects design to assess the categorization of trustworthiness. Consistent with prior experiments, a sample of 86 was needed to detect an interaction between target race and facial trustworthiness. To ensure adequate power, I aimed to oversample and ultimately recruit 100 participants using the same exclusion criteria and sampling method as the previous experiments. However, because of the strike at York University, I recruited only 25 participants from the undergraduate participant pool in exchange for course credit. In an effort to mitigate the effects of the strike, I also recruited participants from the University of Winnipeg's undergraduate participant pool ( $n = 15$ ) for a total of 40 participants. The data from participants who had lived in Canada for less than three years ( $n = 3$ ), who engaged in invariant responding ( $n = 6$ ), and who had incomplete data ( $n = 2$ ) were excluded, leaving a final sample of 29 Black (23 female, 6 male,  $M_{\text{age}} = 20.45$ ,  $SD_{\text{age}} = 4.26$ ) participants. The procedure was identical to that of Experiment 2.

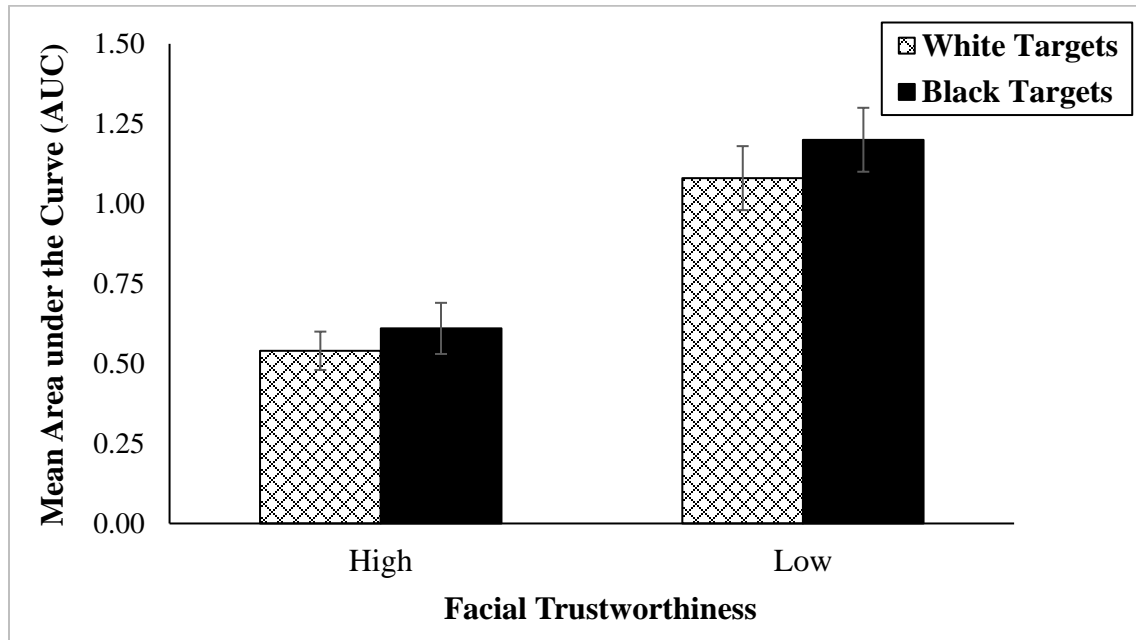


## Results and Discussion

Although it was clear that this experiment was underpowered, I conducted preliminary analyses with the current sample. Specifically, a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) repeated-measures analysis of variance (ANOVA) was performed on participants' AUCs. The main effect of facial trustworthiness was significant,  $F(1, 28) = 11.49, p = .002, \eta_p^2 = 0.29, 95\% \text{ CI } [0.05, 0.52]$ , with larger AUCs for faces low ( $M = 1.14, SD = 0.93$ ) compared to high ( $M = 0.57, SD = 0.47$ ) in facial trustworthiness. The main effect of target race was not significant,  $F(1, 28) = 1.54, p = .225, \eta_p^2 = 0.05, 95\% \text{ CI } [0.00, 0.27]$ , with no difference in AUCs for Black ( $M = 0.90, SD = 0.83$ ) compared to White ( $M = 0.81, SD = 0.75$ ) faces. The two-way interaction was also not significant,  $F(1, 30) = 0.09, p = .761, \eta_p^2 < .01$ , see Figure 9.

**Figure 9**

*Black Participants' Areas under the Curve (AUCs) for Black and White Targets with High and Low Facial Trustworthiness in Experiment 4*



Because of sample size issues, I conducted a post-hoc sensitivity analysis using G\*Power (Faul et al., 2007). This analysis indicated that my final sample ( $N = 29$ ) had .80 power to detect a Facial Trustworthiness  $\times$  Target Race interaction for participants' AUCs of  $F = 2.72$  ( $M$  observed correlation among repeated measures,  $r = .30$ ) with an effect size of  $\eta_p^2 = 0.09$ . Notably, the effect size of the observed interaction,  $\eta_p^2 = .01$ , was much smaller, suggesting that the experiment was underpowered to detect the potential interaction.

The results of Experiment 4, which assessed implicit responses, indicated that Black participants were less direct at categorizing faces low compared to high in facial trustworthiness and that this effect was not qualified by target race. It is important to note, however, that these

findings are only preliminary. Given the small sample size, I aim to recruit more participants in the near future to examine these findings.

### **General Discussion**

Across four experiments, I examined the impact of target and perceiver race on perceptions of facial trustworthiness using both explicit and implicit measures. Although in all experiments, both White and Black perceivers used facial trustworthiness cues to determine perceived trustworthiness on both White and Black faces, the impact of target race on participants' responses differed depending on whether the task was primarily related to implicit or explicit responding and whether the participants were White or Black.

Although White participants in Experiment 1, on an explicit task, rated faces high compared to low in facial trustworthiness as more trustworthy, this effect was qualified by target race. Specifically, participants reported Black compared to White faces as more trustworthy, especially when faces were low in facial trustworthiness. In Experiment 2, White participants demonstrated a different pattern on an implicit mouse-tracking task. In particular, they were less direct at categorizing Black compared to White faces high in facial trustworthiness. Their categorization of faces low in facial trustworthiness, however, was not impacted by target race. Together, these findings demonstrate that although White participants explicitly rated Black compared to White faces depicting low facial trustworthiness as more trustworthy, their responses on a more implicit task suggest a less deliberative tendency to categorize White compared to Black faces depicting high facial trustworthiness as more trustworthy.

Black participants in Experiment 3, on an explicit task, rated both faces high compared to low in facial trustworthiness as more trustworthy and Black compared to White faces as more trustworthy. In Experiment 4, on an implicit mouse-tracking task, participants were less direct at

categorizing faces low compared to high in facial trustworthiness. Notably, target race did not impact the directness of Black participants' mouse trajectories. Together, these findings suggest that although Black participants in general explicitly rate Black compared to White faces as more trustworthy, target race does not impact their less deliberative and implicit responding. It is important to note, however, that the sample size was underpowered in the latter study.

Notably, on the mouse-tracking task, both White and Black participants were less direct in categorizing faces low compared to high in facial trustworthiness. One explanation for this general finding may be that trustworthiness is a more dominant response than untrustworthiness, as previous research has found a dominant response tendency for trust over mistrust that holds across multiple social contexts (Katzir & Posten, 2023). For example, when choosing to trust or mistrust a given target, researchers found that participants' trust compared to mistrust responses were faster. Furthermore, after giving sequential mistrust responses, participants were quicker to switch back to trust responses than vice versa. Given this dominant response tendency, when presented with a face high in facial trustworthiness, participants' initial response tendencies would be to judge it as trustworthy. However, when viewing a face low in facial trustworthiness, participants may need to override this initial tendency. In the present context, that may lead to a larger AUC, with faces low compared to high in facial trustworthiness categorized less directly.

A complementary explanation for the findings on the mouse-tracking task is the negativity bias. Notably, previous research has found that negative stimuli require increased cognitive processing and attention in comparison to positive stimuli (Peeters & Czapinski, 1990), and impression formation and judgement tasks are affected by this bias (Kahneman & Tversky, 1984). Given that people attend more to negative information to arrive at their impressions (Baumeister et al., 2001), less direct categorization of faces low compared to high in facial

trustworthiness in the mouse-tracking task may be related to the more cognitively difficult process of assessing untrustworthy (negative) compared to trustworthy (positive) faces.

### **Social Concerns as a Potential Explanation of Results**

Although social concerns were not directly measured in the current research and further research is required, results from the explicit and implicit tasks suggest that social concerns may have impacted White and Black participants' responses. For White participants, responses on the explicit task provide further support for theorizing that White people are motivated to not express racial bias because of strong social norms (Devine et al., 2002; Kawakami et al., 2017; Shelton, 2000). Although it is possible that White participants rated Black compared to White people as more trustworthy on this task because they genuinely hold this belief, it is notable that this effect was most prominent for faces low in facial trustworthiness. Because evaluating a Black person as untrustworthy may invoke social concerns about appearing prejudiced, White participants may even rate Black faces that do not depict trustworthy features as being relatively trustworthy compared to a White face with similar features. These same concerns would be less evident when rating faces high in facial trustworthiness because both Black and White faces with these types of features would be appropriately rated as trustworthy.

It is also notable that White participants' responses on the explicit and implicit tasks differed. In contrast to their explicit ratings in which they rated Black compared to White faces as more trustworthy, White participants' responses on a more implicit task indicated that their initial tendency was to associate Black faces with untrustworthiness. While the former findings indicate a pro-Black bias and are in accordance with a large literature that social biases and stereotypes may be suppressed because White people are motivated to self-present as unprejudiced (Axt, 2017; Devine et al., 2002), these latter findings are in accordance with a large

literature on more spontaneous negative biases in perceptions of Black people (Brown-Iannuzzi et al., 2019; Hugenberg & Bodenhausen, 2003; Kubota & Ito, 2014; March et al., 2021). In the current mouse-tracking task, more negative responses by White participants may have been related to their inability to monitor prejudice in their responses because they were unaware that their mouse trajectories indicated bias and, therefore, could not adjust their responses according to potential social concerns.

For Black participants, the impact of social concerns was less clear. On their responses to the explicit task, they reported perceiving Black compared to White faces as more trustworthy overall, an effect that was not qualified by the trustworthiness of the facial features of the targets. These findings may be due to Black social concerns related to supporting and positive regard for other Black people (Barr & Neville, 2014; Helms, 1990; Johnson & Carter, 2019; Sellers et al., 1998; Tang et al., 2016), genuinely held positive perceptions of Black compared to White people (Craig & Richeson, 2016; Ho et al., 2017; Roberts et al., 2020; Rotella et al., 2013), or both. Notably, differences in Black participants' responses to the explicit and implicit tasks did not provide clarity on these attributions. In contrast to their explicit ratings, they did not respond differently to Black or White faces on the mouse-tracking task. Although this finding initially suggests that Black participants rate Black compared to White targets more positively when motivated by social concerns and are able to modify their explicit ratings, further research on Black participants' responding on an implicit task is necessary. Given that Experiment 4 was underpowered, I recommend rerunning this study with a larger sample size to properly investigate trustworthiness perceptions using a mouse-tracking task.

Although initial results suggest that social concerns impact participants' explicit perceptions of trustworthiness, future research should directly investigate whether social

concerns moderate the relationship between target race and perceived facial trustworthiness under both implicit and explicit processing conditions. Based on current theorizing, one may expect social concerns to impact explicit but not implicit responses, given that people are both unaware of and less able to control their biases on measures assessing implicit compared to explicit constructs (Dovidio et al., 2009; Nosek, 2007). However, previous research demonstrates that implicit attitudes can be conditioned through socialization (Dovidio et al., 2001; Petty et al., 2006; Wilson et al., 2000), which offers the possibility that cultural socialization related to race may impact both participants' explicit and implicit attitudes in an intergroup context (Tang et al., 2016). In addition, although implicit measures mitigate participants' abilities to control and monitor their responses, it does not fully prevent their concerns from affecting their responses. For example, previous research on non-racial prejudice has found that when completing implicit measures of attitudes toward same-sex attracted targets, participants exhibited lower implicit bias when completing the task in a public compared to a private context (Boysen et al., 2006). It is, therefore, useful to examine whether the influence of these social concerns impact both explicit and implicit processes.

Given that social concerns may differ depending on a participant's race (Richeson & Shelton, 2007), it is important to measure these concerns and how they relate to responses on both explicit and implicit tasks. For White participants, previous scales have been developed that measure motivations to respond without prejudice (Dunton & Fazio, 1997; Plant & Devine, 1998), however, these measures approach prejudiced attitudes more generally and across a wide variety of target groups (Crandall et al., 2002). For Black participants, previous literature has focused primarily on racial identity rather than specific social concerns (Helms, 1990; Sellers et al., 1998).

Therefore, it would be informative for future research to measure social concerns in direct ways that centre both White and Black perspectives in a Black-White intergroup context. For example, based on previous research on intergroup concerns (Richeson & Shelton, 2007), future research could develop measures that directly target concerns of appearing racist toward Black people (for White participants) and supporting other Black people as a result of experienced racism (for Black participants). Notably, White participants with stronger concerns related to appearing racist may be more likely to rate Black and White target faces as equally trustworthy on explicit measures. Similarly, Black participants with stronger concerns related to supporting and positively evaluating Black people may respond similarly to White participants by rating Black faces as highly trustworthy on explicit measures. However, the social concern that moderates this phenomenon may differ and responses on more implicit measures may be telling. Thus, it is necessary to examine how specific social concerns related to both participant and target race affect trustworthiness perceptions on both more deliberate and spontaneous responses.

## **The Impact of Participant Race**

### ***Explicit Perceptions of Trustworthiness***

Although Experiments 1 and 3 were planned and conducted as independent studies, I ran additional analyses to examine whether perceptions of facial trustworthiness statistically differed between Black and White participants on their explicit ratings. Specifically, I combined the datasets from both experiments ( $N = 195$ ) and conducted a 2 (Facial Trustworthiness: High vs. Low) x 2 (Target Race: White vs. Black) x 2 (Participant Race: White vs. Black) mixed analysis of variance (ANOVA) on participants' mean trustworthiness ratings with facial trustworthiness and target race as within-subjects factors and participant race as a between-subjects factor. The



dependent variable was participants' ratings, with higher scores indicating more trustworthiness. The main effect of facial trustworthiness was significant,  $F(1, 193) = 600.87, p < .001, \eta_p^2 = 0.76$ , 95% CI [0.70, 0.80], with participants rating faces high ( $M = 5.55, SD = 1.25$ ) compared to low ( $M = 3.84, SD = 1.23$ ) in facial trustworthiness as more trustworthy. The main effect of target race was also significant,  $F(1, 193) = 154.72, p < .001, \eta_p^2 = 0.44$ , 95% CI [0.35, 0.53], with participants rating Black ( $M = 5.06, SD = 1.46$ ) compared to White ( $M = 4.33, SD = 1.45$ ) faces as more trustworthy. Conversely, the main effect of participant race was not significant,  $F(1, 193) = 0.22, p = .638, \eta_p^2 < .001$ , 95% CI [0.00, 0.03], with no difference in ratings by White ( $M = 4.72, SD = 1.48$ ) compared to Black ( $M = 4.66, SD = 1.53$ ) participants. None of the two-way interactions,  $ps > .150$ , nor the three-way interaction,  $p = .088$ , were significant. Importantly, these results should be interpreted with caution, given that data for both experiments was collected at different time points.

Although these findings suggest, in accordance with previous theorizing, that facial trustworthiness models generalize across perceiver races (Todorov & Oh, 2021; Todorov et al., 2009, 2013), it is notable that when White and Black participants were analyzed separately, two different patterns were found: an interaction effect between facial trustworthiness and target race for White participants (Experiment 1) and only main effects of facial trustworthiness and target race for Black participants (Experiment 3). Given this pattern of findings, I recommend conducting an additional experiment that includes large numbers of both Black and White participants that are run simultaneously.

Notably, when analyzed together, both White and Black participants rated Black targets as more trustworthy. One explanation of this finding may be that racial intergroup interactions present different social concerns for White and Black people that may yield the same result (i.e.,

explicitly rating Black people as more trustworthy; Richeson & Shelton, 2007). While White participants may be concerned about appearing racist, thus rating Black compared to White targets as more trustworthy (Devine et al., 2002; Kawakami et al., 2017; Shelton, 2000), Black participants may be concerned about appearing supportive of other Black people (Helms, 1990; Sellers et al., 1998), which would yield the same pattern.

Additional pilot data conducted from September 2022 to May 2024 supports this theorizing. In this study, when White participants ( $N = 941$ ) were asked to respond to the statement, “In general, I trust Black people will treat me fairly,” on a scale from 1 (strongly disagree) to 5 (neither agree nor disagree) to 9 (strongly agree), the mean response was 6.82 ( $SD = 1.95$ ). Similarly, when asked to respond to the statement, “In general, I trust White people will treat me fairly,” the mean response was 6.86 ( $SD = 1.94$ ). In line with research suggesting that White people are socially concerned with appearing non-racist (Devine et al., 2002; Kawakami et al., 2017, 2019; Shelton, 2000), White participants did not differ in the extent to which they indicated that they would trust White or Black people,  $t(878) = -0.80$ ,  $p = .423$ , 95% CI  $[-0.12, 0.05]$ ,  $d = 0.03$ . Conversely, Black participants’ ( $N = 465$ ) mean response to the statement, “In general, I trust Black people will treat me fairly,” was 6.69 ( $SD = 1.67$ ). However, when asked to respond to the statement, “In general, I trust White people will treat me fairly,” the mean response was 4.70 ( $SD = 1.81$ ). Notably, in accordance with social norms that promote the positive evaluation of and support for Black people (Helms, 1990; Sellers et al., 1998), Black participants trusted that Black compared to White people would treat them more fairly,  $t(452) = 19.38$ ,  $p < .001$ , 95% CI  $[1.80, 2.20]$ ,  $d = 0.91$ .

### ***Implicit Perceptions of Trustworthiness***

Although Experiments 2 and 4 were planned and conducted as independent studies, I ran additional analyses to examine whether implicit responses on a mouse-tracking task differed between Black and White participants. Specifically, I combined datasets from both experiments ( $N = 124$ ) and conducted a 2 (Facial Trustworthiness: High vs. Low)  $\times$  2 (Target Race: White vs. Black)  $\times$  2 (Participant Race: White vs. Black) mixed analysis of variance (ANOVA) on participants' mean trustworthiness ratings with facial trustworthiness and target race as within-subjects factors and participant race as a between-subjects factor. The dependent variable was participants' AUCs, with a larger AUC (i.e., less direct response trajectory) indicating a stronger association with the unchosen category (Freeman et al., 2008). The main effect of facial trustworthiness was significant,  $F(1, 115) = 22.42, p < .001, \eta_p^2 = 0.16, 95\% \text{ CI } [0.06, 0.29]$  with larger AUCs for faces low ( $M = 1.23, SD = 0.97$ ) compared to high ( $M = 0.71, SD = 0.65$ ) in facial trustworthiness. The main effect of target race was close to significance,  $F(115) = 3.95, p = .049, \eta_p^2 = 0.03, 95\% \text{ CI } [0.00, 0.12]$ , with larger AUCs for Black ( $M = 1.00, SD = 0.88$ ) compared to White faces ( $M = 0.93, SD = 0.85$ ). The main effect of participant race, however, was not significant,  $F(1, 115) = 2.12, p = .148, \eta_p^2 = 0.02, 95\% \text{ CI } [0.00, 0.09]$ , with no differences between AUCs for Black ( $M = 0.86, SD = 0.79$ ) and White ( $M = 1.00, SD = 0.88$ ) participants. The two-way interactions,  $ps > .317$ , and the three-way interaction,  $p = .161$ , were not significant. Notably, these results should be interpreted with extreme caution, given that data for both experiments was collected at different time points, and the study focusing on Black participants was underpowered.

Given that when White and Black participants were analyzed separately, two separate patterns were found—an interaction effect between facial trustworthiness and target race for White participants (Experiment 2) and a main effect of facial trustworthiness but not of target

race for Black participants (Experiment 4)—future research should investigate whether perceiver race impacts participants' implicit responses with a larger sample of both White and Black participants.

## **Methodological Implications**

### ***Real and Artificial Faces***

It is important to note that the current research used artificial (i.e., computer-generated) rather than real faces and that researchers have found an increased cognitive burden for faces that are not photorealistic (Albohn et al., 2022). For example, previous research has identified a number of features on artificial faces that make them more difficult to process and recall (Balas & Pacella, 2017; Gaither et al., 2019), such as a lack of skin pigmentation and low contrast in the eyes (Balas & Tonsager, 2014). Importantly, other research has demonstrated that perceptions of trustworthiness can differ between artificial and real faces (Balas & Pacella, 2017). In particular, artificial faces are perceived as less trustworthy, and participants tend to be more successful in identifying whether faces are high or low in facial trustworthiness when they are real rather than artificial.

Although participants in the current research were able to perceive differences between high and low trustworthiness on artificial faces, future research should employ images of real people to increase the ecological validity of the current pattern of results. Although I was able to standardize the stimuli in the current research by carefully controlling how their features varied across different levels of facial trustworthiness for both Black and White targets, these systematic differences may not reflect variation in real-world faces. For example, a real face with large, round eyes (a feature contributing to perceptions of high trustworthiness) may also have

low inner eyebrows (a feature contributing to perceptions of low trustworthiness; Todorov & Oosterhof, 2011).

To examine how such feature combinations impact impressions, in future research, participants could rate the objective presence of a number of facial trustworthiness features on a set of images of real White and Black faces. For example, in a first phase, participants could be asked to assess how large and round a target face's eyes are or how low a set of eyebrows is. Given findings in the current research that target race impacts perceptions of trustworthiness, participants could also be asked to assess race-related items such as the racial prototypicality of facial features. Based on these feature ratings, it would be possible to average responses across participants and create an aggregate score for both the facial trustworthiness (e.g., low, medium, or high) and racial prototypicality (e.g., low, medium, or high) of each target stimuli. In a second phase, a group of naïve participants could rate how much they trust each target face. It would then be possible to examine whether trustworthiness perceptions from the second phase correlated with the objective ratings provided in the first phase. This design would allow the researcher to pinpoint the impact of the interaction between target and perceiver race on trustworthiness perceptions and provide a more nuanced understanding of how facial cues and target race affect trustworthiness perceptions in a more realistic context.

### ***Racial Diversity***

While previous research on facial trustworthiness has conventionally employed White target faces (Collova et al., 2019; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Swe et al., 2020; Todorov et al., 2009, 2013; Vernon et al., 2014; Willis & Todorov, 2006), the current research utilized stimuli of Black and White faces as well as Black and White perceivers. This is notable, given that first impressions can be influenced by both bottom-up cues related to a

target's facial features and top-down characteristics related to perceiver motivations (Freeman & Ambady, 2011; Hehman et al., 2017; Kawakami et al., 2017). Future research, however, could benefit from considering other participant and target racial groups, given that perceptual processes and evaluations can be impacted by perceivers' lived experiences and cultural beliefs about particular target groups (Bijlstra et al., 2014; Hugenberg & Bodenhausen, 2003; Kang & Chasteen, 2009; Mende-Siedlecki et al., 2019; Roberts et al., 2020; Xie et al., 2021).

For example, given that the current research demonstrates that White people (a racial majority group) likely compensate for their prejudices by rating Black targets (a racial minority group) as highly trustworthy, it is important to examine whether these effects generalize to other racial majority-minority relationships, such as White perceivers and East Asian targets. For White perceivers, considerations of the specific histories of North American racism against Black people (e.g., slavery, segregation; Roberts et al., 2020) may uniquely impact their trustworthiness perceptions. For example, given the concerns around appearing racist toward a historically marginalized racial group, White people may possess strong tendencies to explicitly rate Black people as highly trustworthy, while the strength of societal-level racism toward Black people may strengthen White people's implicit associations between Black people and untrustworthiness. In contrast, although other target racial groups such as East Asians have their own specific histories of oppression and discrimination (Chen et al., 2020; Lu et al., 2020; Sue et al., 2007), these targets may potentially elicit a different response from White participants as Black and East Asian groups are generally assigned different perceptions and stereotypes (Fiske et al., 2002).

Black perceivers' specific histories may also uniquely impact their trustworthiness perceptions in a Black-White intergroup context. For example, results from the current research

demonstrate that when Black perceivers (a racial minority group) are asked to explicitly report facial trustworthiness, they rate White targets (a racial majority group) as less trustworthy than Black targets regardless of facial trustworthiness levels. An important consideration is whether these effects generalize to other intergroup relationships, such as a minority-minority relationship like that between Black perceivers and East Asian targets. Notably, Black people's specific lived and historical experiences of being the target of racism (Roberts et al., 2020) suggest that their negative expectations as to how White people will treat them can negatively impact their perceptions of how trustworthy White people are (Brown & Dobbins, 2004; Cohen & Steele, 2002). Conversely, although previous research demonstrates that an emphasis on intergroup differences can sometimes encourage negative relationships between minority groups (Bastian & Haslam, 2008; Chen & Ratliff, 2018), inter-minority coalitions can also be strengthened by emphasizing common identities (Dovidio et al., 2009; Ho et al., 2017) and experiences of collective victimization (Cortland et al., 2017; Craig & Richeson, 2012; Rotella et al., 2013). Thus, in contrast to Black people's negative perceptions of White targets' trustworthiness, their perceptions of racial minority groups such as East Asians may be comparatively more positive.

Future research should, therefore, focus on other races and consider the specific nuances of different intergroup relationships. One strategy would be to utilize existing diverse face databases to source real images (e.g., Chen et al., 2020; Lakshmi et al., 2020; Ma et al., 2015; Strohminger et al., 2016) and use facial manipulation software (Singular Inversions, 2024) to transform racialized faces along a continuum of facial trustworthiness (Todorov et al., 2011, 2013). Such stimuli would allow for further investigation of how perceptions of trustworthiness are impacted by the interaction between perceiver and target race outside of the Black-White intergroup context.

### *Gender Diversity*

One methodological limitation of the current research is that it was constrained by the use of exclusively male faces, a bias that was purposefully introduced in the original stimuli set because male compared to female faces with no hair were perceived as more natural (Oosterhof & Todorov, 2008). Although this bias allowed for the ability to standardize stimuli so that any effects could be attributed to the perception of facial features and not peripheral cues such as hairstyle, there may be differences in trustworthiness perceptions based on gender. For example, previous research indicates that female faces may be perceived as more trustworthy than male faces (Luo et al., 2023; Walker & Wänke, 2017; Wirth & Wentura, 2023), and even male faces with more feminine than unfeminine features may be perceived as more trustworthy (Oosterhof & Todorov, 2008).

In addition, participants may perceive White and Black faces in a more intersectional way when considering gender. For example, researchers have found that participants exhibit neural pattern similarities during early face processing of seemingly unrelated gender and race categories, such as Black with male and Asian with female (Stolier & Freeman, 2016). In the case of facial trustworthiness, it is possible that such intersectional categories could bias perceptions. For example, given that female faces have a perceived trustworthiness advantage (Luo et al., 2023; Walker & Wänke, 2017; Wirth & Wentura, 2023), Asian faces may also benefit from this advantage due to their neural overlap with female faces. Furthermore, previous research has found that people process target race as early as 122 ms after stimulus presentation but target sex 50 ms later (Ito & Urland, 2003). Therefore, it may be informative to examine how these two cues interact when forming rapid, first impressions. For example, would White people perceive Black women, Black men, White women, or White men as more trustworthy, and



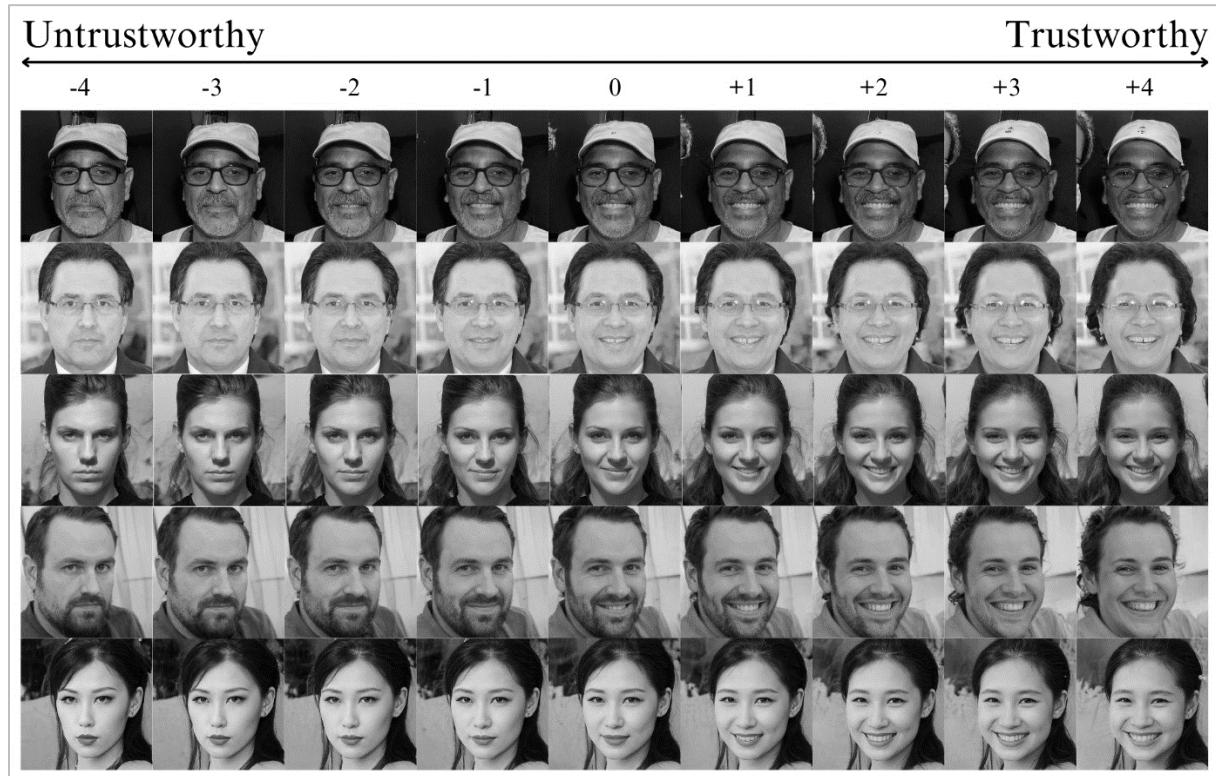
would this perception change depend on whether responses were on an explicit versus implicit task?

### ***Machine Learning***

Although, in comparison to real faces, people often rate traditional computer-generated faces differently (Balas & Pacella, 2017; Gaither et al., 2019), new methods using machine learning have produced artificial faces indistinguishable from real ones. Notably, research has shown that when presented with both real and machine-learning-synthesized faces, participants are unable to correctly classify the images as real or artificial above chance (Nightingale & Farid, 2021). In addition, recent research has successfully applied such techniques to generate large numbers of photorealistic faces upon which researchers can model various gradations of trait perceptions including trustworthiness (Albohn et al., 2022; Peterson et al., 2022). For example, Figure 10 contains a sample of realistic, machine-learning-generated faces created with Extempo (Uddenberg et al., 2024), a generative artificial intelligence that models its face judgements based on previous human perceptions (Albohn et al., 2022).

**Figure 10**

*Examples of Machine-Learning Generated Faces Modelling Perceived Trustworthiness*



*Note.* Each row represents one target face, while each column represents a single directional vector from -4 to +4 in perceived trustworthiness.

Given the methodological importance of extending the current research to real faces, more racially diverse faces, and more gender-diverse faces, the application of machine-learning techniques may provide a useful avenue to increase the ecological validity of this research. First, given that facial trustworthiness models do not always neatly map onto the features of real faces, machine-learning techniques may be beneficial to examine how these effects generalize in a more controlled format while still maintaining the realism of real faces. Second, the use of machine learning can reduce the difficulties in sourcing stimuli from underrepresented groups,

which are often limited in face stimuli databases (Olivola et al., 2014). Unlike real-life stimuli, machine-learning methods allow the researcher to control both demographic and trait-judgement concerns while still maintaining a high level of realism. For example, Figure 10 captures how machine-learning technologies can generate stimuli across a variety of target races (e.g., Black, White, East Asian) and genders (e.g., men and women) that vary along a continuum of perceived trustworthiness. Overall, future research would benefit from using such methods to examine perceptions of trustworthiness that capture ecologically valid and diverse stimuli.

## **Conclusion**

The ability to understand when another person is trustworthy is foundational to all relationships (Simpson, 2007), yet it is particularly critical between racial groups whose relations may be, at times, fraught with suspicion (Brewer, 1999; Dovidio et al., 2002; Kramer & Messick, 1998; Lloyd et al., 2017; Tropp, 2008). The current research suggests that perceptions of facial trustworthiness, while generalizable at a universal level, are also impacted by both the race of the perceiver and the target. Furthermore, whether these perceptions are measured using an explicit or implicit measure matters. Although White participants explicitly reported perceiving Black relative to White targets as highly trustworthy (particularly in conditions where they were low in facial trustworthiness), their responses on an implicit task revealed that they held associations between Black faces and untrustworthiness. Conversely, Black participants explicitly reported perceiving Black relative to White targets as highly trustworthy, yet in an underpowered implicit task, they demonstrated no association between target race and trustworthiness. Although the current research measured perceptions of trustworthiness in explicit and implicit laboratory tasks, it has the potential to shed light on the importance of how target and perceiver race can interact in real-world contexts. One such context is when White perceptions of (un)trustworthiness result

in the disproportionate racial bias that Black people face as a result of systemic, perceived suspicion that they have committed crimes even when no criminal activity has occurred (Griggs, 2018; Pierson et al., 2020).

“I’m not a racist,” Amy Cooper—a White woman from New York City—protested in 2020, even after her behaviour in Central Park suggested otherwise (Vera & Ly, 2020). Despite her explicit insistence that there were “never any racial implications” to her words (Cooper, 2023), in an encounter captured on video, Amy was pictured growing increasingly agitated and untrusting of the video’s recorder while she dialed 911, saying, “I’m going to tell them there’s an African American man threatening my life.” The “African American man” and video-recorder in question—who is referred to in these racialized terms three times across a one-minute video—was Christian Cooper (no relation), a Black man who was birdwatching nearby and became concerned that Amy’s off-leash dog might be affecting the habitat of the ground-dwelling birds in the area. This incident—only one example on a long list of mundane activities done by Black people that are perceived as untrustworthy or suspicious by White people and have resulted in calls to the police (Griggs, 2018)—illustrates a similar pattern to the current research in which, despite explicitly reporting positive perceptions of Black people, White people display implicit racial bias and trust them less than they are willing to admit.

Although this is only one example of an interaction between a White perceiver and Black target, such an incident illustrates the importance of the current research and how the race of the perceiver and target can impact how perceptions of trustworthiness unfold in a particular encounter. Given the different patterns of findings for White and Black perceivers and targets in the current research, it is likely that if Amy had been Black and Christian had been White, or if both people were White or both were Black, Amy’s perceptions may have been different.

Importantly, this real-world incident highlights how both explicit and implicit perceptions contribute to such situations, in addition to demonstrating that social concerns related to perceiver and target race may impact perceptions differently.

Ultimately, by demonstrating that the relationship between the race of the perceiver and target can affect processes of perceiving trustworthiness, these findings extend previous literature that has focused primarily on the impact of facial feature cues alone on assessments of trustworthiness. Although future research is necessary to further examine these issues, these initial findings suggest that perceiver and target race do matter and understanding how they influence trait attributions is not only integral to improving intergroup relations but also to understanding the universality of face perception processes.

## References

- Adams, R. B., Jr., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, 5(1), 3–11.  
<https://doi.org/10.1037/1528-3542.5.1.3>
- Albohn, D. N., Uddenberg, S., & Todorov, A. (2022). A data-driven, hyper-realistic method for visualizing individual mental representations of faces. *Frontiers in Psychology*, 13.  
<https://doi.org/10.3389/fpsyg.2022.997498>
- Axt, J. R. (2018). The best way to measure explicit racial attitudes is to ask about them. *Social Psychological & Personality Science*, 9(8), 896–906.  
<https://doi.org/10.1177/19485506177289>
- Balas, B., & Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Computers in Human Behavior*, 77, 240–248. <https://doi.org/10.1016/j.chb.2017.08.045>
- Balas, B., & Tonsager, C. (2014). Face animacy is not all in the eyes: Evidence from contrast chimeras. *Perception*, 43(5), 355–367. <https://doi.org/10.1068/p7696>
- Barr, S. C., & Neville, H. A. (2014). Racial socialization, color-blind racial ideology, and mental health among Black college students: An examination of an ecological model. *Journal of Black Psychology*, 40(2), 138–165. <https://doi.org/10.1177/0095798412475084>
- Bastian, B., & Haslam, N. (2008). Immigration from the perspective of hosts and immigrants: Roles of psychological essentialism and social identity. *Asian Journal of Social Psychology*, 11(2), 127–140. <https://doi.org/10.1111/j.1467-839X.2008.00250.x>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>

- Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, 46(4), 657–663. <https://doi.org/10.1016/j.jesp.2010.03.006>
- Boysen, G. A., Vogel, D. L., & Madon, S. (2006). A public versus private administration of the implicit association test. *European Journal of Social Psychology*, 36(6), 845–856. <https://doi.org/10.1002/ejsp.318>
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, 55(3), 429–444. <https://doi.org/10.1111/0022-4537.00126>
- Brown-Iannuzzi, J. L., Cooley, E., McKee, S. E., & Hyden, C. (2019). Wealthy Whites and poor Blacks: Implicit associations between racial groups and wealth predict explicit opposition toward helping the poor. *Journal of Experimental Social Psychology*, 82, 26–34. <https://doi.org/10.1016/j.jesp.2018.11.006>
- Brown, L. M., & Dobbins, H. (2004). Students' of color and European American students' stigma-relevant perceptions of university instructors. *Journal of Social Issues*, 60(1), 157–174. <https://doi.org/10.1111/j.0022-4537.2004.00104.x>
- Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, 61(2), 87–105. <https://doi.org/10.1016/j.cogpsych.2010.03.001>
- Chen, J. A., Zhang, E., & Liu, C. H. (2020). Potential impact of COVID-19-related racial discrimination on the health of Asian Americans. *American Journal of Public Health*, 110(11), 1624–1627. <https://doi.org/10.2105/AJPH.2020.305858>
- Chen, J. M., & Ratliff, K. (2018). Psychological essentialism predicts intergroup bias. *Social Cognition*, 36(3), 301–323. <http://dx.doi.org/10.1521/soco.2018.36.3.301>

- Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*, 88(6), 885–894. <https://doi.org/10.1037/0022-3514.88.6.885>
- Cohen, G. L., & Steele, C. M. (2002). A barrier of mistrust: How negative stereotypes affect cross-race mentoring. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 303–327). Academic Press.  
<https://doi.org/10.1016/B978-012064455-1/50018-X>
- Collova, J. R., Sutherland, C. A. M., & Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children’s faces are not the same as for adults’ faces. *Journal of Personality and Social Psychology*, 117(5), 900–924.  
<https://doi.org/10.1037/pspa0000167>
- Cooper, A. (2023, November 28). I was branded the ‘Central Park Karen’. I still live in hiding. *Newsweek*. <https://www.newsweek.com/i-was-branded-central-park-karen-i-still-live-hiding-1839483>
- Cortland, C. I., Craig, M. A., Shapiro, J. R., Richeson, J. A., Neel, R., & Goldstein, N. J. (2017). Solidarity through shared disadvantage: Highlighting shared experiences of discrimination improves relations between stigmatized groups. *Journal of Personality and Social Psychology*, 113(4), 547–567. <https://doi.org/10.1037/pspi0000100>
- Craig, M. A., & Richeson, J. A. (2016). Stigma-based solidarity: Understanding the psychological foundations of conflict and coalition among members of different stigmatized groups. *Current Directions in Psychological Science*, 25(1), 21–27.  
<https://doi.org/10.1177/0963721415611252>



- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–377. <https://doi.org/10.1037//0022-3514.82.3.359>
- Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology*, 8(2), 88–102. <https://doi.org/10.1037//1099-9809.8.2.88>
- Dovidio, J. F., Kawakami, K., Smoak, N., & Gaertner, S. L. (2009). The roles of implicit and explicit processes in contemporary prejudice. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 165–192). Psychology Press.
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, 82(5), 835–848. <https://doi.org/10.1037/0022-3514.82.5.835>
- Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8), 2455–2483. <https://doi.org/10.1093/rfs/hhs071>
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23(3), 316–326. <https://doi.org/10.1177/0146167297233009>
- 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>

- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902.  
<https://doi.org/10.1037/0022-3514.82.6.878>
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, 20(10), 1183–1188.  
<https://doi.org/10.1111/j.14679280.2009.02422.x>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241. <https://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of personal construal. *Psychological Review*, 118(2), 247–279. <https://doi.org/10.1037/a0022327>
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, 137, 673–690. <https://doi.org/10.1037/a0013875>
- Friesen, J., Kawakami, K., Vingilis-Jaremko, L., Caprara, R., Memmo, J., Meyers, C., & Fang, X. (under review). Smiling while Black: Trustworthiness judgments of true and false smiles on Black and White faces. *Journal of Experimental Psychology: General*.
- Friesen, J., & Sinclair, L. (2011). Distrust and simultaneous activation of multiple categories. *Social Psychological & Personality Science*, 2(1), 112–118.  
<https://doi.org/10.1177/1948550610382666>
- Fruhen, L. S., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks.

*The Leadership Quarterly*, 26(6), 1005–1016.

<https://doi.org/10.1016/j.leaqua.2015.07.001>

Funder, D., Levine, J., Mackie, D., Morf, C., Sansone, C., Vazire, S., & West, S. (2014).

Improving the dependability of research in personality and social psychology:

Recommendations for research and educational practice. *Personality and Social*

*Psychology Review*, 18(1), 3–12. <https://doi.org/10.1177/1088868313507536>

Gaither, S. E., Chen, J. M., Pauker, K., & Sommers, S. R. (2019). At face value: Psychological

outcomes differ for real vs. computer-generated multiracial faces. *The Journal of Social*

*Psychology*, 159(5), 592–610. <https://doi.org/10.1080/00224545.2018.1538929>

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in

evaluation: An integrative review of implicit and explicit attitude change. *Psychological*

*Bulletin*, 132(5), 692–731. <https://doi.org/10.1037/0033-2909.132.5.692>

Griggs, B. (2018, December 28). Living while Black. *CNN*.

<https://www.cnn.com/2018/12/20/us/living-while-black-police-calls-trnd/index.html>

Helman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique

contributions of perceiver and target characteristics in person perception. *Journal of*

*Personality and Social Psychology*, 113(4), 513–529.

<https://doi.org/10.1037/pspa0000090>

Helms, J. E. (1990). *Black and White racial identity: Theory, research, and practice*. Greenwood

Press.

Ho, A. K., Kteily, N. S., & Chen, J. M. (2017). “You’re one of us”: Black Americans’ use of

hypodescent and its association with egalitarianism. *Journal of Personality and Social*

*Psychology*, 113(5), 753–768. <https://doi.org/10.1037/pspi0000107>

- Holoien, D. S., Bergsieker, H. B., Shelton, J. N., & Alegre, J. M. (2015). Do you really understand? Achieving accuracy in interracial relationships. *Journal of Personality and Social Psychology*, 108(1), 76–92. <https://doi.org/10.1037/pspi0000003>
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, 14(6), 640–643.  
[https://doi.org/10.1046/j.0956-7976.2003.psci\\_1478.x](https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x)
- Hugenberg, K., & Wilson, J. P. (2013). Faces are central to social cognition. In D. E. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 167–193). Oxford University Press.
- Hutchings, R. J., Freiburger, E., Sim, M., & Hugenburg, K. (2024). Racial prejudice affects representations of facial trustworthiness. *Psychological Science*, 35(3), 263–276.  
<https://doi.org/10.1177/09567976231225094>
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85(4), 616–626. <https://doi.org/10.1037/0022-3514.85.4.616>
- Jaeger, B., & Jones, A. L. (2021). Which facial features are central in impression formation? *Social Psychological and Personality Science*, 13(2), 553–561.  
<https://doi.org/10.1177/19485506211034979>
- Johnson, V. E., & Carter, R. T. (2019). Black cultural strengths and psychosocial well-being: An empirical analysis with Black American adults. *Journal of Black Psychology*, 46(1), 55–89. <https://doi.org/10.1177/0095798419889752>
- Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo,

- M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., ... Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, 5(1), 159–169.  
<https://doi.org/10.1038/s41562-020-01007-2>
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *The American Psychologist*, 39(4), 341–250. <https://doi.org/10.1037/0003-066X.39.4.341>
- Kang, S. K., & Chasteen, A. L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *International Journal of Aging & Human Development*, 69(3), 201–220. <https://doi.org/10.2190/AG.69.3.c>
- Karmali, F., & Kawakami, K. (2023). Posing while black: The impact of race and expansive poses on trait attributions, professional evaluations, and interpersonal relations. *Journal of Personality and Social Psychology*, 124(1), 49–68.  
<https://doi.org/10.1037/pspa0000313>
- Karmali, F., Kawakami, K., Vaccarino, E., Williams, A., Phills, C., & Friesen, J. P. (2019). I don't see race (or conflict): Strategic descriptions of ambiguous negative intergroup contexts. *Journal of Social Issues*, 75(4), 1002–1034. <https://doi.org/10.1111/josi.12353>
- Katzir, M., & Posten, A. C. (2023). Are there dominant response tendencies for social reactions? Trust trumps mistrust-evidence from a Dominant Behavior Measure (DBM). *Journal of Personality and Social Psychology*, 125(1), 57–81. <https://doi.org/10.1037/pspa0000334>
- Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology* (pp. 1–80). Elsevier Academic Press. <https://doi.org/10.1016/bs.aesp.2016.10.001>

- Kawakami, K., Hugenberg, K., & Dunham, Y. (2020). Perceiving others as group members: Basic principles of social categorization processes. In P. van Lange, E. T. Higgins, and A. W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles*, (3rd ed., pp. 411–429). Guilford Press.
- Kawakami, K., Meyers, C., & Friesen, J., (in press). Impact of true and false smiles on perceived trustworthiness of Black and White faces: A Black perspective. Manuscript submitted to *Journal of Experimental Psychology: General*.
- Kramer, R. M., & Messick, D. M. (1998). Getting by with a little help from our enemies: Collective paranoia and its role in intergroup relations. In C. Sedikides (Ed.), *Intergroup Cognition and Intergroup Behavior* (pp. 233–255). Lawrence Erlbaum Associates.
- Kubota, J. T., & Ito, T. A. (2014). The role of expression and race in weapons identification. *Emotion*, 14(6), 1115–1124. <https://doi.org/10.1037/a0038214>
- Lakshmi, A., Wittenbrink, B., Correll, J., & Ma, D. S. (2021). The India face set: International and cultural boundaries impact face impressions and perceptions of category membership. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.627678>
- Lloyd, E. P., Kunstman, J. W., Tuscherer, T., & Bernstein, M. J. (2017). The face of suspicion: Suspicion of Whites' motives moderates mental representations of Whites. *Social Psychological and Personality Science*, 8(8), 953–960. <https://doi.org/10.1177/1948550617699251>
- Lo, R. F., & Sasaki, J. Y. (2024). Lay misperceptions of culture as “biological” and suggestions for reducing them. *Perspectives on Psychological Science*, 19(1), 295–311. <https://doi.org/10.1177/17456916231181139>

- Lu, J., Nisbett, R. E., & Morris, M. W. (2020). Why East Asians but not South Asians are underrepresented in leadership positions in the United States. *PNAS*, *117*(9), 4590–4600. <https://doi.org/10.1073/pnas.1918896117>
- Luo, X., Song, J., Guan, J., & Chen, L. (2024). Influence of facial dimorphism on interpersonal trust: weighing warmth and competence traits in different trust situations. *Current Psychology*, *43*, 2158–2172 <https://doi.org/10.1007/s12144-023-04472-w>
- 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>
- March, D. S., Gaertner, L., & Olson, M. A. (2021). Danger or dislike: Distinguishing threat from negative valence as sources of automatic anti-Black bias. *Journal of Personality and Social Psychology*, *121*(5), 984–1004. <https://doi.org/10.1037/pspa0000288>
- Mende-Siedlecki, P., Qu-Lee, J., Backer, R., & Van Bavel, J. J. (2019). Perceptual contributions to racial bias in pain recognition. *Journal of Experimental Psychology: General*, *148*(5), 863–889. <https://doi.org/10.1037/xge0000600>
- Nosek, B. A. (2007). Understanding the individual implicitly and explicitly. *International Journal of Psychology*, *42*(3), 184–188. <https://doi.org/10.1080/00207590601068>
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 556–570. <https://doi.org/10.1016/j.tics.2014.09.007>
- Oosterhof, N. N., & Todorov, A. (2008). Functional basis of face evaluation. *PNAS*, *105*(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>

- Oswald, F., & Adams, R. B. (2023). Feminist social vision: Seeing through the lens of marginalized perceivers. *Personality and Social Psychology Review*, 27(3), 332–356. <https://doi.org/10.1177/10888683221126582>
- Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology*, 1(1), 33–60. <https://doi.org/10.1080/14792779108401856>
- Peterson, J. C., Uddenberg, S., Griffiths, T. L., & Suchow, J. W. (2022). Deep models of superficial face judgments. *PNAS*, 119(17). <https://doi.org/10.1073/pnas.2115228119>
- Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST Model. *Journal of Personality and Social Psychology*, 90(1), 21–41. <https://doi.org/10.1037/0022-3514.90.1.21>
- Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behavior*, 4, 736–745. <https://doi.org/10.1038/s41562-020-0858-1>
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832. <https://doi.org/10.1037/0022-3514.75.3.811>
- 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>



- Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, 16(6), 316–320. <https://doi.org/10.1111/j.1467-8721.2007.00528.x>
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, 15(6), 1295–1309. <https://doi.org/10.1177/1745691620927709>
- Roberts, S. O., & Rizzo, M. T. (2021). The psychology of American racism. *American Psychologist*, 76(3), 475–487. <https://doi.org/10.1037/amp0000642>
- Rotella, K. N., Richeson, J. A., Chiao, J. Y., & Bean, M. G. (2013). Blinding trust: The effect of perceived group victimhood on intergroup trust. *Personality and Social Psychology Bulletin*, 39(1), 115–127. <https://doi.org/10.1177/0146167212466114>
- Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, 125(2), 207–218. <https://doi.org/10.1016/j.cognition.2012.06.017>
- Sellers, R. M., Smith, M. A., Shelton, J. N., Rowley, S. A. J., & Chavous, T. M. (1998). Multidimensional model of racial identity: A reconceptualization of African American racial identity. *Personality and Social Psychology Review*, 2(1), 18–39. [https://doi.org/10.1207/s15327957pspr0201\\_2](https://doi.org/10.1207/s15327957pspr0201_2)
- Shelton, J. N. (2000). A reconceptualization of how we study issues of racial prejudice. *Personality and Social Psychology Review*, 4(4), 374–390. [https://doi.org/10.1207/S15327957PSPR0404\\_6](https://doi.org/10.1207/S15327957PSPR0404_6)
- Slepian, M. L., Chun, J. S., & Mason, M. F. (2017). The experience of secrecy. *Journal of Personality and Social Psychology*, 113(1), 1–33. <https://doi.org/10.1037/pspa0000085>

- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354.  
<https://doi.org/10.1177/0956797610362647>
- Strohming, 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. <https://doi.org/10.3758/s13428-015-0641-9>
- Stolier, R., & Freeman, J. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*, 795–797. <https://doi.org/10.1038/nn.4296>
- Simpson, J. A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *16*(5), 264–268.  
<https://doi.org/10.1111/j.1467-8721.2007.00517.x>
- Singular Inversions. (2024). *FaceGen Modeller* (3.22). <https://facegen.com>
- Sue, D. W., Bucceri, J., Lin, A. I., Nadal, K. L., & Torino, G. C. (2007). Racial microaggressions and the Asian American experience. *Asian American Journal of Psychology*, *5*(1), 88–101. <https://doi.org/10.1037/1948-1985.S.1.88>
- Sutherland, C. A. M., Burton, N. S., Wilmer, J. B., Blokland, G. A. M., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *PNAS*, *117*(19), 10218–10224.  
<https://doi.org/10.1073/pnas.1920131117>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118.  
<https://doi.org/10.1016/j.cognition.2012.12.001>

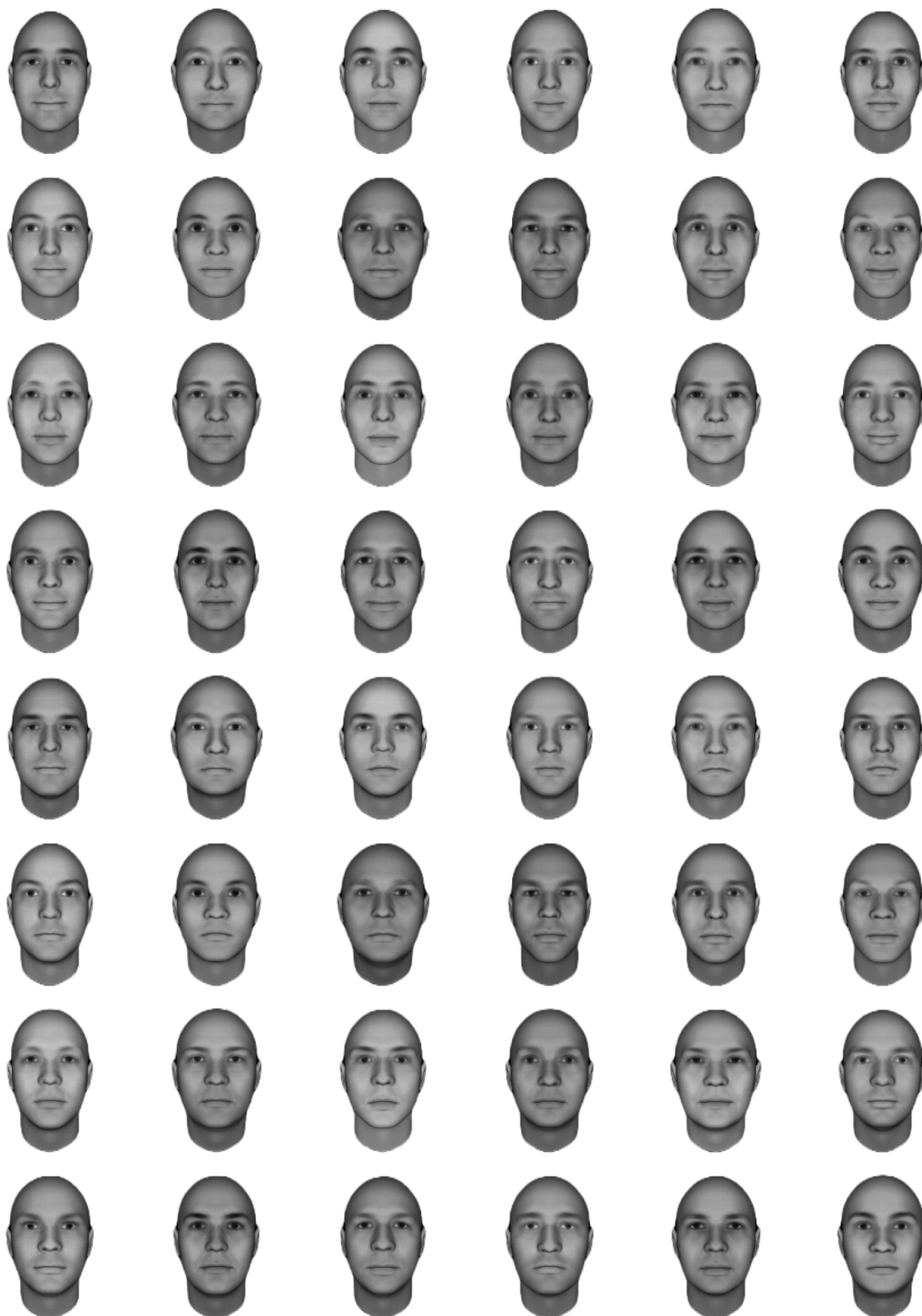
- Swe, D. C., Palermo, R., Gwinn, O. S., Rhodes, G., Neumann, M., Payart, S., & Sutherland, C. A. M. (2020). An objective and reliable electrophysiological marker for implicit trustworthiness perception. *Social Cognitive and Affective Neuroscience*, *15*(3), 337–346. <https://doi.org/10.1093/scan/nsaa043>
- Swim, J. K., Cohen, L. L., & Hyers, L. L. (1998). Experiencing everyday prejudice and discrimination. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The Target's Perspective* (pp. 37–60). Academic Press. <https://doi.org/10.1016/B978-012679130-3/50037-5>
- Tang, S., McLoyd, V. C., & Hallman, S. K. (2016). Racial socialization, racial identity, and academic attitudes among African American adolescents: Examining the moderating influence of parent–adolescent communication. *Journal of Youth and Adolescence*, *45*(6), 1141–1155. <https://doi.org/10.1007/s10964-015-0351-8>
- Thierry, S. M., & Mondloch, C. J. (2021). First impressions of child faces: Facial trustworthiness influences adults' interpretations of children's behaviour in ambiguous situations. *Journal of Experimental Child Psychology*, *208*. <https://doi.org/10.1016/j.jecp.2021.105153>
- Todorov, A. (2017). *Face Value: The Irresistible Influence of First Impressions*. Princeton University Press. <https://doi.org/10.1515/9781400885725>
- Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–738. <https://doi.org/10.1037/a0032335>
- Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, *5*(10), 775–791. <https://doi.org/10.1111/j.1751-9004.2011.00389.x>

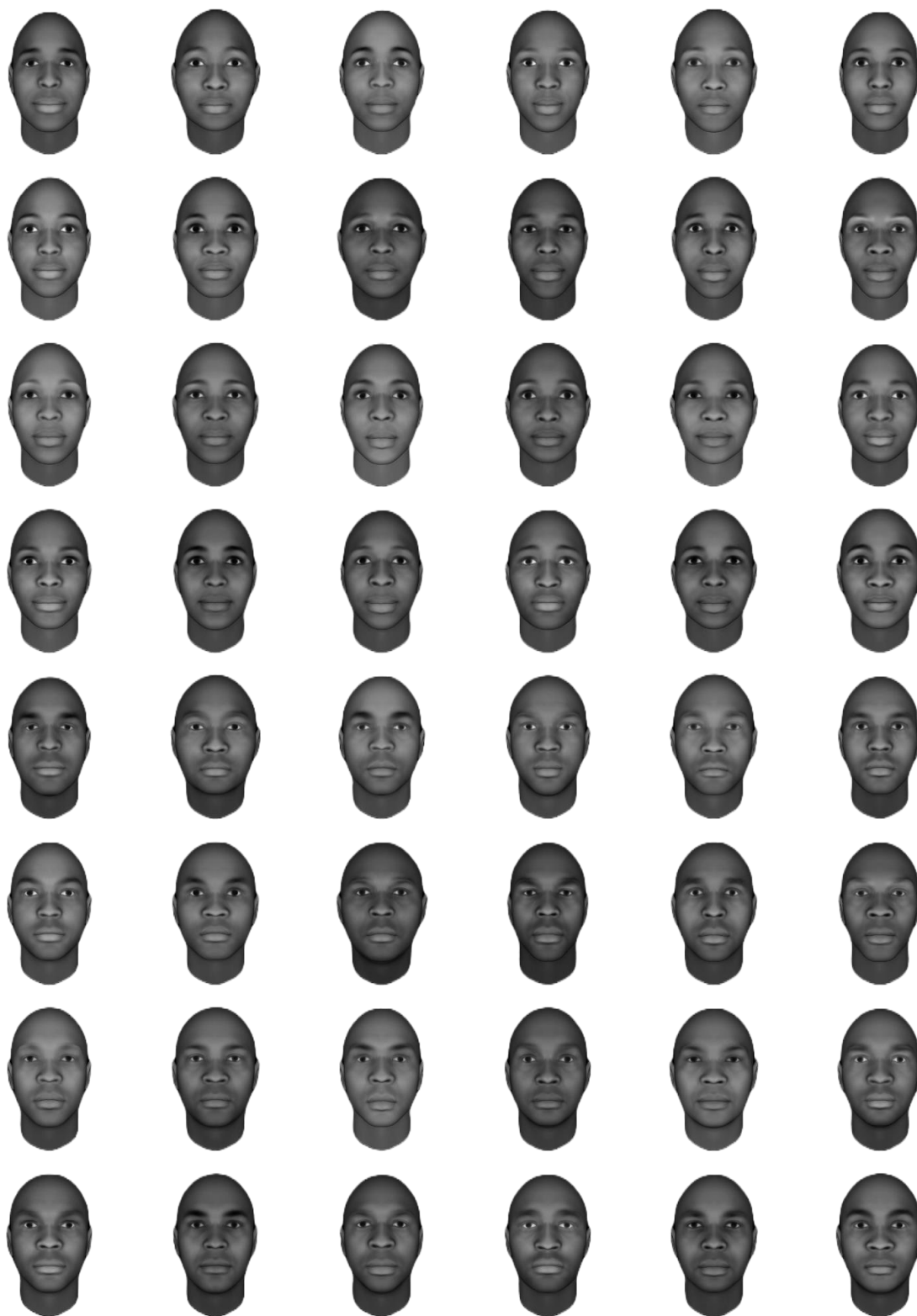
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, 28(2), 117–122. <https://doi.org/10.1109/MSP.2010.940006>
- 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>
- Tropp, L. R. (2008). The role of trust in intergroup contact: Its significance and implications for improving relations between groups. In U. Wagner, L. R. Tropp, G. Finchilescu, & C. Tredoux (Eds.), *Improving intergroup relations: Building on the legacy of Thomas F. Pettigrew* (pp. 91–106). Blackwell Publishing.  
<https://doi.org/10.1002/9781444303117.ch7>
- Uddenberg, S., Shah, R., & Albohn, D. (2024). *Extempo*. <https://extempo.rocks>
- Vera, A., & Ly, L. (2020, May 26). White woman who called police on a Black man bird-watching in Central Park has been fired. *CNN*.  
<https://www.cnn.com/2020/05/26/us/central-park-video-dog-video-african-american-trnd/index.html>

- Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, *111*(32), E3353–E3361.  
<https://doi.org/10.1073/pnas.140986011>
- Vorauer, J. D. (2005). Miscommunications surrounding efforts to reach out across group boundaries. *Personality and Social Psychology Bulletin*, *31*(12), 1653–1664.  
<https://doi.org/10.1177/0146167205277808>
- Walker, M., & Wänke, M. (2017). Caring or daring? Exploring the impact of facial masculinity/femininity and gender category information on first impressions. *PloS one*, *12*(10). <https://doi.org/10.1371/journal.pone.0181306>
- 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>
- Wirth, B. E., & Wentura, D. (2023). Not lie detection but stereotypes: Response priming reveals a gender bias in facial trustworthiness evaluations. *Journal of Experimental Social Psychology*, *104*, 104406-. <https://doi.org/10.1016/j.jesp.2022.104406>
- Xie, S. Y., Flake, J. K., Stoller, R. M., Freeman, J. B., & Hehman, E. (2021). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science*, *32*(12), 1979–1993. <https://doi.org/10.1177/09567976211024259>

## Appendix A

### Stimuli





## **Appendix B**

### Demographic Questions

We ask for information about your general background so that we can accurately describe our participant sample. The information you provide below will be kept confidential and not used to identify you.

1. With what gender do you most identify?
2. What is your age in years?
3. What is your ethnicity/cultural background? You may choose as many categories as you identify with.
4. How many years have you lived in Canada?
5. How many years have you been speaking English?
6. Because this study asks you to respond to images, we are interested in if any of the following statements apply to you. You can select as many statements as apply to you.
  - a. I have been formally diagnosed with an attentional disorder (e.g., ADHD). If yes, please indicate which one(s).
  - b. I have been formally diagnosed with a mood disorder (e.g., depression, bipolar disorder). If yes, please indicate which one(s).
  - c. I have been formally diagnosed with an anxiety disorder (e.g., generalized anxiety disorder). If yes, please indicate which one(s).
  - d. None of these statements apply to me.



## Appendix C

### Study Experience Questions

Your answers on these final questions will not affect your compensation so please feel free to respond openly and honestly. We will use your responses to improve future surveys.

1. This survey will be used to answer important research questions. If you were tired or distracted when you were answering the questions, or didn't read the instructions carefully, your responses may not be accurate. Were you able to stay focused throughout the study? (Please be honest; there's no penalty if you answer 'no.')
2. Have you seen any of the images used in this study before today? If so, where have you seen them before?
3. What do you think the purpose of the experiment was?
4. What do you think the experimenter expected to find?
5. Do you have any comments on what it was like for you to complete this study (e.g., were any parts particularly confusing, interesting, boring or awkward)?
6. Did you take any breaks while completing the study?
7. While completing the study, was there anything that affected your responses on this task?  
If so, please describe.