

**“Other”:  
Examining the Link Between Race-Based Biases in Social Cognition and Social  
Perception**

by  
Natalie Medico

A thesis submitted to the Faculty of the University of Delaware in partial  
fulfillment of the requirements for the degree of Bachelor of Arts in Psychology with  
Distinction

Fall 2023


© 2023 Natalie Medico  
All Rights Reserved


**“Other”:**


**Examining the Link Between Race-Based Biases in Social Cognition and Social Perception**

by

Natalie Medico

Approved:   
Peter Mende-Siedlecki, Ph.D.  
Professor in charge of thesis on behalf of the Advisory Committee

Approved:   
Jasmin Cloutier, Ph.D.  
Committee member from the Department of Psychological and Brain Sciences

Approved:   
Jaclyn Schwarz, Ph.D.  
Committee member from the Board of Senior Thesis Readers

Approved: \_\_\_\_\_  
Dana Veron, Ph.D.  
Chair of the University Committee on Student and Faculty Honors

## **ACKNOWLEDGMENTS**

I would like to give my greatest thanks my thesis director, Dr. Peter Mende-Siedlecki. This thesis would not have been possible without his continued support, both throughout this process and during the overall two years I have worked alongside him. I would also like to thank the other members of my thesis board, Dr. Jasmin Cloutier and Dr. Jaclyn Schwarz, for their assistance on this project. Finally, I would like to acknowledge both my former lab manager, Patrick Reyes, and my closest friends and family for motivating me the past few years. I would not be where I am today without them.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	iii
LIST OF FIGURES .....	vi
ABSTRACT .....	vii
1 INTRODUCTION .....	1
a) Intergroup bias .....	1
b) Intergroup bias, race, and social perception .....	2
c) Intergroup bias, race, and impression formation .....	4
d) The present research .....	6
2 EXPERIMENT 1 .....	8
a) Methods .....	8
i) Participants .....	8
b) Materials .....	8
i) Face stimuli .....	9
ii) Behavior stimuli .....	9
iii) Procedure .....	10
iv) Analyses .....	12
c) Results .....	12
i) Cross-Race-Effect .....	12
ii) Initial Impression Task .....	13
iii) Updating Task .....	14
(1) Effects of target race on impression updating .....	15

(2) Correlational analyses .....	16
d) Discussion.....	16
3 EXPERIMENT 2.....	18
a) Methods .....	18
i) Participants .....	18
b) Materials .....	19
i) Face stimuli .....	19
ii) Behavior stimuli .....	19
iii) Procedure .....	19
iv) Analyses .....	21
c) Results .....	21
i) Initial impression task.....	21
ii) Updating task.....	22
(1) Four-way interaction between initial valence, updated valence, time, and race.....	23
(2) Four-way interaction between initial valence, updated valence, time, and pain condition .....	25
iii) Pain perception task.....	28
iv) Correlational analyses .....	29
d) Discussion.....	30
4 GENERAL DISCUSSION .....	31
REFERENCES .....	38
APPENDIX .....	43

## LIST OF FIGURES

Figure 1: Effects of target race on face sensitivity .....	13
Figure 2: Effects of behavior valence and target race on initial impression formation (Experiment 1).....	14
Figure 3: Effects of behavior valence and target race on initial impressions (Experiment 2).....	22
Figure 4: Effects of target race and valence on impression updating.....	25
Figure 5: Effects of behavior valence and pain condition on updating magnitude ....	27
Figure 6: Effects of behavior valence and target race on pain perception.....	29

## ABSTRACT

Sociality is inherent to human life. Although the processes that underlie it often help us navigate the world around us efficiently, they are often subject to bias, namely ingroup bias. Moreover, biases in perception (e.g., the Cross-Race-Effect) may be related to biases in our judgments of others (i.e., impression formation). The present research aims to understand the dynamic nature of racial biases by testing if differences in impression formation and updating are correlated with differential perception of faces. More specifically, we hypothesized that Black individuals would be rated lower on trustworthiness based on learned behavioral information, and that these ratings would be correlated with blunted sensitivity to Black faces during memorization and pain tasks. 270 participants were recruited over two experiments; in the first experiment, subjects first completed a standard Cross-Race-Effect (CRE) task. They then learned either positive or negative false information about either Black or White targets, after which they made initial impression ratings. Participants then learned new information about the target and were prompted to update their impressions. In Experiment 2, participants completed the same Impression Formation and Impression Updating tasks as Experiment 1, but a Pain Perception Task was interlaced between ratings (as opposed to the CRE task). Results from both experiments suggest that race and valence can impact impression formation and updating, and that differential sensitivity to faces may be correlated with differences in impressions under certain conditions. This research presents new findings on the interconnected nature of racial bias and provides new insight on the potential causes of everyday microaggressions that many people of color experience.

## **Chapter 1**

### **INTRODUCTION**

#### **Intergroup bias**

Sociality is an inherent aspect of human life. The majority of our day-to-day existence is taken up with thinking about, interacting with, and looking at other humans. It has been argued that our contemporary social behavior has its origins in survival tactics that provided a series of evolutionary benefits (Dunbar, 2011). In other words, sociality increased survival, and thus, this adaptation has remained a core component of humanity as a result. Although seemingly complex, much of this social behavior builds upon our ability to engage in automatic processing of others (e.g., Bargh, 2013; Gilbert, 1989; Winter & Uleman, 1984), allowing us to rapidly make judgements that dictate our social responses. Though these processes help us navigate social decisions quickly and efficiently, they can be subject to bias—in particular, biases associated with group membership.

Research on intergroup bias suggests that we may automatically evaluate and perceive members of ingroup and outgroups differently, which in turn can facilitate differences in behavior towards those who are labeled as “other” (e.g., Brewer, 1979; Brewer, 2010). In other words, our tendency to categorize individuals can lead to preferential treatment of those with whom we share group membership, known as ingroup favoritism. This effect can manifest itself in several ways, such as more favorable outcomes for ingroup members (e.g., bigger rewards), greater protection of the ingroup, and better resource allocation for the ingroup (Schiller et al.,



2013; Tajfel et al., 1971). Closely related to ingroup favoritism is outgroup derogation, or the active dislike of or disdain for outgroup members. Said derogation can lead to harsher treatment, such as more severe punishments and less pardoning, less empathetic responses, or on a more extreme level, aggressive and violent behavior (Cikara et al., 2011; Molenberghs, 2013). In sum, individuals work to protect others with whom they share group membership, either by uplifting those close to them or by reprimanding those who are not.

### **Intergroup bias, race, and social perception**

While the extensive body of research on intergroup bias has shown that these group assignments can be made based on a series of dimensions – including group divisions that are arbitrary or minimal in nature (Tajfel et al., 1971; Van Bavel, Packer, & Cunningham, 2008) – the lessons of this literature can be clearly applied to race-based conflict. As in these other instances, individuals often perceive their racial ingroup more positively than their outgroup (Dovidio et al., 1997). As a result of this ingroup favoritism, racial outgroup members are often treated more poorly than their ingroup counterparts. Additionally, explicit or implicit racial attitudes may be amplified by racial stereotypes that have been systematically used to maintain racial divisions and hierarchy. Thus, modern prejudicial attitudes towards outgroup members (and resulting discriminatory behavior) may be related to the racial application of ingroup favoritism coupled with outgroup derogation (Greenwald & Pettigrew, 2014).

As with other group-based divisions, the effects of intergroup bias can lead to differences in behavior. Trust decisions, as an example, can be made based on predisposed beliefs of an outgroup when objective, factual information is unavailable.

Implicit racial attitudes, then, can lead to differences in perceived trustworthiness based on an individual's race, as one may use negative racial stereotypes as an aid in decision making when other diagnostic information is omitted (Stanely et al, 2011; 2012). This effect can also be seen beyond trust decisions, however. In instances where an action is performed by a racially ambiguous person, individuals may display stereotype-congruent (e.g., prejudicial) behavior when evaluating the morality of said individual, as diagnostic information is limited (Devine, 1989). Thus, ingroup favoritism can lead to the formation of negative stereotypes against a racial outgroup, which can shape social decision-making and consequent behavior.

Divergent behavior towards members of different racial groups isn't confined to direct social interaction, however. Intergroup bias may also be reflected in differential perceptions of in-group and out-group members. An extensive literature demonstrates that we are worse at processing the faces of racial out-group members (Meissner & Brigham, 2001). For example, White participants struggle to individuate, or differentiate, other-race faces (demonstrating the "outgroup homogeneity effect;" Hughes et al., 2019, see also Ackerman et al., 2006 and Corneille et al., 2007) and show worse memory for Black and Asian faces compared to White faces (often referred to as the "Cross Race Effect" [CRE]; (Meissner & Brigham, 2001; Young, Hugenberg, & Sacco, 2012). These biases may be linked to a tendency to process same-race faces configurally (i.e., in terms of the second-order relationships between aspects of the face), while other-race face processing is more featural in nature (Hancock & Rhodes, 2008). This differential engagement of face processing mechanisms may have demonstrable consequences for behavior, particularly in the context of eyewitness identification (Wilson, Hugenberg, & Bernstein, 2013).

Moreover, these differences in basic face processing are paralleled by and may lead to differences in emotion recognition. For example, recent work shows that White participants have stricter thresholds for seeing pain on Black faces, which predicted biases in hypothetical treatment recommendations (Mende-Siedlecki et al., 2019; 2021; 2022; Xu et al., 2009). Like the CRE, this perceptual bias is also linked to differences in configural face processing, specifically that other-race faces are often processed featurally. This bias can be observed in Experiments 3-4 of work by Mende-Siedlecki and colleagues (2019), where participants saw Black and White faces in different presentation orientations. In the condition where faces were presented in an upright orientation, racial bias in pain perception was maintained. However, when faces were in an inverted orientation (such that configural processing is disrupted, but featural processing is maintained), this anti-Black bias was reduced. Based on these results, the researchers concluded that perceptual bias in pain recognition was associated, in part, with reduced configural processing of Black faces. Thus, perceivers' engagement of basic face perception mechanisms may vary when viewing members of a racial outgroup compared to a racial ingroup, and in turn, this divergence can be reliably indexed in standard behavioral paradigms. Additionally, disruptions in configural face processing can have consequences for emotion recognition and humanization (see also Cassidy et al., 2017 and Deska & Hugenberg, 2017), and may fuel discriminatory behavior towards marginalized groups (Fincher & Tetlock, 2016), Black individuals in particular.

### **Intergroup bias, race, and impression formation**

A second aspect of sociality that is impacted by intergroup dynamics is impression formation and updating. Humans form strong links between other people

(Todorov & Uleman, 2002; Bliss-Moreau et al., 2008; Todorov & Olson, 2008), automatically inferring character traits from their behaviors to support social impressions (Todorov & Uleman, 2003). In some cases, we are quicker to form lasting impressions based on stereotypes, as opposed to individuating information, but these judgements can be swayed by the perceiver's own personal beliefs, motivations, and general views of outgroup members (Kunda et al., 1996). For example, while negative or immoral information about another person's character is generally seen as more diagnostic for our impressions (Mende-Siedlecki et al., 2013; 2016), these patterns may shift once group membership is accounted for. For example, other work shows that perceivers differentially rely on novel positive and negative information when learning about ingroup versus outgroup targets, respectively (Hughes et al., 2016). This tendency to align our social impressions with group-based expectations may also be reflected in social cognitive research examining how target race moderates impression formation. For example, EEG research demonstrates that White perceivers rely on automatic processes when forming judgements of Black and White targets based on stereotypes, but that reaction times are slower when these targets are presented with expectancy-violating information (Dickter et al., 2012). Similar fMRI work has also demonstrated that affective associations guide impression formation and observed that brain regions involved in mentalizing (e.g., dorsomedial prefrontal cortex) are preferentially activated when viewing information that violates racial stereotypes (Li et al., 2016), as well as when individuating same-race (but not other-race) individuals (Freeman et al., 2010).

As alluded to above, impressions do not always remain fixed. When our initial impressions are violated or contradicted, we must update these impressions based on

the newly learned information. However, several studies have shown that impression updating is often asymmetrical. Within the domain of morality, in particular, participants show greater updates of their impressions when learning new negative information about a target, compared to new positive information (Mende-Siedlecki et al., 2013; 2016; Kim et al., 2020). Based on this work, the magnitude of impression updating is dictated by directionality of expectation violations (positive to negative or vice versa). However, other factors – in particular, whether the expectancy violation was done by a racial ingroup or outgroup member – may affect the magnitude of updating, too. In particular, racial stereotypes can guide impression formation, and the difficulty to disregard these generalized beliefs may maintain negative impressions of racial outgroup members, and consequently, fuel behaviors that uphold systemic racism.

### **The present research**

Ingroup favoritism and outgroup derogation can, respectively, motivate individuals to maintain positive views of fellow ingroup members but apply negative stereotypes to members of the outgroup. If, however, these stereotypes modulate impression formation processes, then it is possible that racial ingroup favoritism can modulate asymmetrical impression updating. That is, White perceivers may be quicker to form more positive initial impressions of White individuals and negative initial impressions of Black individuals and may be more resistant to impression updating of racial outgroup members in the face of new, contradictory information. Additionally, although extensive work has been done to identify the perceptual and cognitive bases of racial bias, little has been done to explore the potential relationship between these two sources. The tendency for perceivers to individuate ingroup members, but to

generalize outgroup members (e.g., Hugenberg et al., 2012), may have consequences beyond social perception (for example, reduced recognition of out-group identity and emotion) and social cognition (for example, blunted updating in response to out-group behavior).

Thus, the present research aims to address the following three questions: (1) Do subjects form more negative initial impressions of other-race faces compared to same-race faces? (2) Do subjects show reductions in impression updating when learning about other-race-faces compared to same-race faces? (3) Do race-based differences in impression updating predict racial bias in social perceptual tasks assessing pain perception and face memory? Specifically, we predicted that White participants would form more negative impressions of, and show reduced impression updating for, Black faces. Moreover, we theorized that these biases would be associated with worse encoding and emotion perception when considering Black faces. Specifically, we predicted that differential impression formation and updating would be correlated with racial bias tasks assessing cross-race face memory and pain perception.

## Chapter 2

### EXPERIMENT 1

To begin with, we examined the relationship between group-based bias in impression formation and updating and concordant bias in face memory. Participants learned sets of valenced behavioral information about Black and White targets, which in some cases switched valence half-way through the set. Thus, this task provided indices of both initial impressions and impression updates. In a separate task, participants were asked to memorize the faces of a different set of Black and White targets and, after a delay, were tested on their recognition memory of these individuals. We predicted that tendencies to show more negative impressions of and reduced updates for Black (versus White) targets would be correlated with worse memory for Black (versus White) targets in the face memory task.

#### Methods

**Participants.** 140 participants ( $M_{\text{age}} = 18.43$  years,  $SD = 0.72$ ; 119 female, 19 male, 2 non-binary/gender-queer; 88% White, 3% Black) were recruited from the University of Delaware via SONA, an online recruitment platform used to credit students enrolled in PSYC100 for research participation. Since recruitment for Experiment 1 and Experiment 2 overlapped, we ensured that no participants could take part in both studies. All participants completed all aspects of each task, therefore no recruited participants needed to be excluded from data analysis.

#### Materials

**Face stimuli.** We employed a standard Cross Race Effect (CRE) task paradigm for measuring group-based differences in face memory. We used 120 images (60 Black male faces, 60 White male faces, all neutral expressions) adapted from previous work (Eberhardt et al., 2001; Hughes et al., 2019). Three additional image pairs were also used for the distractor task between the Learning and Test phases of the CRE task. We selected three images of outdoor settings, where the second image was edited so small objects were removed.

For the Impression Formation and Updating tasks, we selected 24 targets (12 Black, 12 White) from the Delaware Pain Database (DPD; Mende-Siedlecki et al., 2020) making neutral facial expressions. These images were presented on a transparent background and were resized to 300 pixels by 300 pixels.

**Behavior stimuli.** 288 behavioral statements (144 positive, 144 negative) were selected from the Delaware Behavior Database (DBD; Mende-Siedlecki & Havlicek, in prep). To select these specific statements, we pared down the larger set of 2375 behaviors by applying criteria to ensure that our selections were sufficiently “charged” in terms of valence, yet not too *extreme* in their content. Specifically, negative statements scoring lower than -3 on valence and higher than 5.25 on arousal were removed, meanwhile positive statements scoring higher than +3 on valence and 5.25 on arousal were removed. Statements were then filtered further to minimize the differences between the negative and positive statement sets in terms of absolute valence ( $M_{\text{Negative}} = 2.030$ ,  $SD_{\text{Negative}} = 0.478$ ,  $M_{\text{Positive}} = 2.143$ ,  $SD_{\text{Negative}} = 0.592$ ;  $p = .076$ ), moral relevance ( $M_{\text{Negative}} = 5.304$ ,  $SD_{\text{Negative}} = 0.774$ ,  $M_{\text{Positive}} = 5.213$ ,  $SD_{\text{Negative}} = 0.902$ ;  $p = .355$ ), arousal ( $M_{\text{Negative}} = 3.392$ ,  $SD_{\text{Negative}} = 0.622$ ,  $M_{\text{Positive}} = 3.544$ ,  $SD_{\text{Negative}} = 0.660$ ;  $p = .047$ ), and frequency ( $M_{\text{Negative}} = 20.466$ ,  $SD_{\text{Negative}} = 9.179$ ,



$M_{\text{Positive}} = 20.860$ ,  $SD_{\text{Negative}} = 6.570$ ;  $p = .676$ ) (Note that while the positive behaviors we selected were ultimately slightly higher in terms of absolute valence and arousal, if anything, this difference would run counter to predictions based on the extensive person perception literature suggesting that negative (i.e., immoral) behaviors are more diagnostic when forming and updating impressions.) Critically, the negative behavior set was seen as considerably less trustworthy than the positive behavior set ( $M_{\text{Negative}} = -1.727$ ,  $SD_{\text{Negative}} = 0.666$ ,  $M_{\text{Positive}} = 1.615$ ,  $SD_{\text{Negative}} = 0.473$ ;  $p = 2.541 \times 10^{-141}$ )

**Procedure.** After reading and agreeing to the digital consent form, participants were instructed on the experimental tasks. Presentation order of the CRE task and the Impression Formation and Updating task was counterbalanced. All portions of the experimental paradigm were built and presented using the Qualtrics survey platform.

The CRE task was directly adapted from prior work (Hugenberg et al., 2007). For participants who began with the CRE task, they first completed the *learning phase*. During this phase, participants viewed 20 Black and 20 White neutral faces that they were asked to memorize. Each face was shown for three seconds.

Following previous CRE paradigms, participants then completed a *distractor phase*, where they were shown three sets of images, where one image was slightly different from the other. Subjects were instructed to look for as many differences between the images as possible. A text box appeared for one minute on screen, in which subjects could write down the differences they found. The screen auto-advanced after one minute.

After the *distractor phase*, participants underwent the *testing phase*, where they were once again shown 80 Black and White neutral faces. 40 of the images were

shown during the learning phase, while the other 40 images were “new” to the subject. Participants were instructed to look at each image and indicate if it was “old” (i.e., shown during the learning phase) or “new” (i.e., *not* shown during the learning phase). This portion of the task was self-paced.

In addition to completing the CRE task, participants completed Impression Formation and Updating tasks using a separate set of Black and White targets. On each trial, participants began with the Impression Formation task, where they viewed a neutral Black or White male face paired with either a positive or negative behavioral statement revealing something about that individual’s moral character. Each statement was presented to the participant for five seconds before auto-advancing to the next. After reading four behavioral statements, participants were asked to rate how positively they felt about the individual based on trustworthiness, using a scale from -4 to 4 (-4 = very untrustworthy; 4 = very trustworthy).

Next, participants then completed the Impression Updating task, where they viewed the same neutral stimulus from before, but paired with two new morality statements. After learning this new information, subjects were asked to rate, once again, how trustworthy they thought the individual was on a scale from -4 to 4. Trustworthiness ratings in both the Impression Formation and Updating task were self-paced.

As the experiment was counterbalanced, the presentation order of the tasks was randomized, so that certain participants (N = 72) completed the CRE task first, while the rest (N = 68) began with the Impression tasks. All participants learned about 24 individuals in the Impression Formation and Updating tasks, each comprising 6 behaviors.

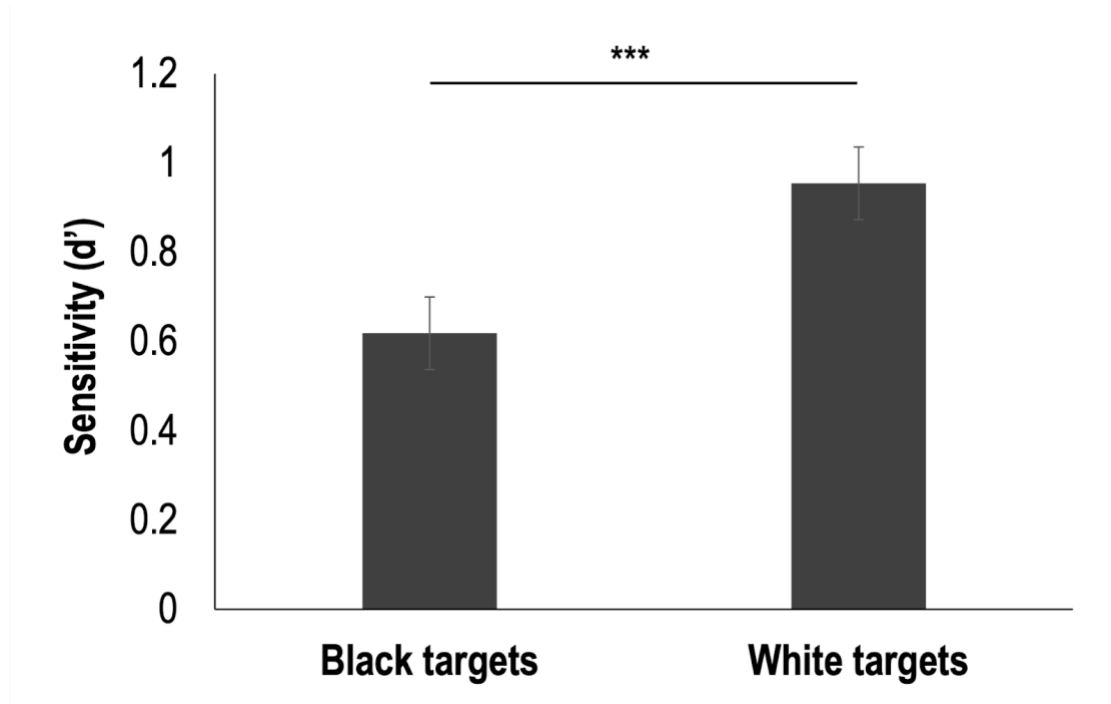
Subjects were finally asked to complete a demographic questionnaire, as well as a series of questions on personal beliefs about “Americanness” related to another ongoing experiment being conducted in our lab. Participation in this latter questionnaire wasn’t obligatory.

**Analyses.** Our analytic plan comprised three successive steps. First, we conducted a paired *t*-test to test whether participants’ sensitivity for recognizing targets in the CRE task varied by target race. Second, we conducted a 2 (valence: positive vs. negative)  $\times$  2 (target race: Black vs. White) repeated measures ANOVA to test if the effect of valence on initial impression formation varied by target race. Third, we conducted a 2 (target race: Black vs. White)  $\times$  2 (initial valence: positive vs. negative)  $\times$  2 (updated valence: positive vs. negative)  $\times$  2 (time: first rating vs. second rating) repeated measures ANOVA to see if the effects of updating condition on updating magnitude varied by target race.

Subsequently, we tested to see if racial bias in initial impression formation, face memory, and updating magnitude were correlated with each other.

## **Results**

**Cross-Race Effect task.** We observed a significant main effect of target race on participants’ *d'* scores in the CRE task ( $t(139) = 5.681, p < .001, d = 0.480$ ). Specifically, participants showed greater sensitivity when making Old/New judgments of White faces during the test phase ( $M = 0.954, SD = 0.730$ ), versus Black faces ( $M = 0.618, SD = 0.599$ ) (See Figure 1).

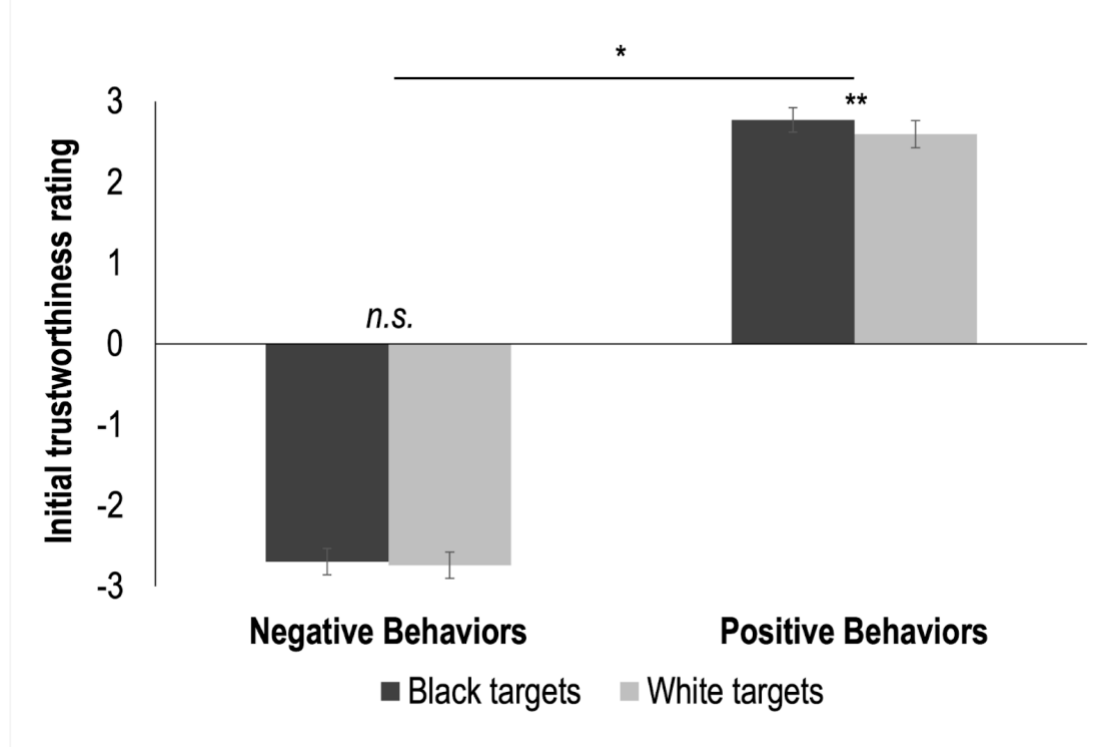


**Figure 1. Effects of target race on face sensitivity.** We observed a significant effect of race on memorization and recall of faces, such that participants showed greater sensitivity of White faces compared to Black. Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008). <sup>†</sup> $p < .10$ , <sup>ns</sup> $p > .10$

**Initial impression task.** We observed a strong, statistically significant main effect of behavior valence on initial impression formation ( $F(1,139) = 2177.184$ ,  $p < .001$ ,  $\eta_p^2 = 0.940$ ). As would be expected, participants rated targets associated with negatively-valenced behaviors as much less trustworthy ( $M = -2.778$ ,  $SD = 0.793$ ) than targets associated with positively-valenced behaviors ( $M = 2.632$ ,  $SD = 0.847$ ).

While the main effect of target race was not statistically significant ( $F(1,139) = 2.223$ ,  $p = .138$ ,  $\eta_p^2 = 0.16$ ), the interaction between target race and valence was marginally significant ( $F(1,139) = 3.473$ ,  $p = .064$ ,  $\eta_p^2 = 0.024$ ). To better understand this effect, we ran separate  $t$ -tests comparing Black and White targets at either level of initial valence. Within targets initially paired with negative behaviors, we observed no difference in trustworthiness ratings as a function of race ( $t(139) = 0.116$ ,  $p = .908$ ,  $d =$

0.010,  $M_{\text{Black}} = -2.781$ ,  $SD_{\text{Black}} = 0.881$ ;  $M_{\text{White}} = -2.773$ ,  $SD_{\text{White}} = 0.89$ ). However, within targets initially paired with positive behaviors, Black targets received significantly higher trustworthiness ratings than their White counterparts ( $t(139) = 2.720$ ,  $p = .007$ ,  $d = 0.230$ ;  $M_{\text{Black}} = 2.702$ ,  $SD_{\text{Black}} = 0.88$ ;  $M_{\text{White}} = 2.561$ ,  $SD_{\text{White}} = 0.911$ ) (See Figure 2).



**Figure 2. Effects of behavior valence and target race on initial impression formation (Experiment 1).** A main effect of behavior valence on impression formation was observed. While we did not find a significant overall effect of target race on initial impressions, Black targets who were paired with initially positive behaviors were rated higher on trustworthiness (compared to White). Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008).  $^{\dagger}p < .10$ ,  $^{n.s.}p > .10$

**Updating task.** As predicted, we observed a significant three-way interaction between initial behavioral valence, updated behavioral valence, and time ( $F(1,139) = 90.934$ ,  $p < .001$ ,  $\eta_p^2 = 0.395$ ). To break down this three-way interaction, we ran separate two-way interactions examining the effects of updated valence and time at either level of initial valence, collapsing across race.

Within targets whose initial behavior was positive, this two-way interaction between updated valence and time was statistically significant ( $F(1,139) = 829.857, p < .001, \eta_p^2 = 0.857$ ). Breaking this down further, we observed a small, but significant rise in trustworthiness ratings for targets whose behavior was consistently positive across the two rating time points ( $t(139) = 6.554, p < .001, d = 0.554, M = -0.284, SD = 0.513$ ), while we observed a large, statistically significant drop in trustworthiness ratings for targets whose behavior changed from positive to negative valence ( $t(139) = -23.471, p < .001, d = -1.984, M = 2.98, SD = 1.503$ ).

Within targets whose initial behavior was negative, this two-way interaction between updated valence and time was also statistically significant ( $F(1,139) = 299.326, p < .001, \eta_p^2 = 0.683$ ). In subsequent analyses, we observed a small, but significant drop in trustworthiness ratings for targets whose behavior was negative across the two ratings ( $t(139) = -3.327, p < .001, d = -0.281, M = 0.188, SD = 0.667$ ), while we observed a large, statistically significant rise in trustworthiness ratings for targets whose behavior valence changed from negative to positive ( $t(139) = 14.008, p < .001, d = 1.184, M = -1.927, SD = 1.628$ ). Notably, the change in trustworthiness ratings was larger going from positive to negative compared to negative to positive, replicating previous findings in the updating literature.

***Effects of target race on impression updating.*** Despite our predictions, the four-way interaction that included race was not statistically significant ( $F(1,139) = 0.223, p = .432, \eta_p^2 = 0.004$ ). The updating effect (i.e., Initial  $\times$  Updating  $\times$  Time interaction) was consistent within both Black ( $F(1,139) = 48.791, p < .001, \eta_p^2 = 0.260$ ) and White targets ( $F(1,139) = 57.910, p < .001, \eta_p^2 = 0.294$ ). Indeed, no other

interactions in this study involving race were statistically significant (all  $ps > .156$  [race  $\times$  initial valence  $\times$  time interaction]).

Even the main effect of target race itself was only marginally significant ( $F(1,139) = 3.610, p = .060, \eta_p^2 = 0.025$ ). Specifically, participants tended to rate White targets ( $M = -0.230, SD = 0.544$ ) as slightly more *untrustworthy* than Black targets ( $M = -0.155, SD = 0.532$ ), counter to our expectations.

***Correlational analyses.*** We assessed the correlational relationships between bias in judgments in the CRE task (i.e.,  $d'$  for White targets minus  $d'$  for Black targets) and four separate biases in the Impression Formation and Updating tasks. Participants' bias in the CRE task was positively and significantly correlated with a tendency to show greater impression updates for White (versus Black) targets whose behavior changed in valence from negative to positive ( $r(140) = 0.224, p = .008$ ). However, we saw no relationship between bias in the CRE task and biases in initial impressions, overall updating magnitude, or updating specifically in the positive to negative condition (all  $rs < 0.084$ , all  $ps > .322$ ).

## **Discussion**

Taken together, we replicated previous findings suggesting that a) White perceivers show worse memory for Black (versus White) faces in the CRE paradigm and b) in general, negative (i.e., immoral) behaviors produce larger impression updates than positive (i.e., moral) behaviors (even when minimizing baseline differences between these behaviors as a function of valence). That said, participants did not show differences as a function of target race in terms of their initial impressions or impression updates. (If anything, participants rated Black targets associated with positive behaviors as being more trustworthy than their White counterparts.) However,

with regards to our main research question, one notable relationship across tasks was observed: to the extent that participants showed greater sensitivity when making Old/New judgments of White (versus Black) faces in the CRE paradigm, they also tended to show greater improvements in their impressions of White (vs. Black) targets in the negative-to-positive condition.



### Chapter 3

## EXPERIMENT 2

Moving forward, we examined the relationship between group-based bias in impression formation and updating and concordant bias in emotion perception. Participants completed the same Impression Formation and Updating tasks as in Experiment 1, but these were now separated by a pain perception task for each individual target. Here, participants saw a set of morph images ranging from neutral to painful expressions and rated the intensity of each individual morph. This task allows us to determine participants' thresholds for perceiving pain on Black and White faces. We predicted that tendencies to show more negative impressions of and reduced updates for Black (versus White) targets would be correlated with a tendency to see pain less readily on the faces of Black (versus White) targets in the pain perception task.

Notably, while outside the scope of our primary aims, the structure of this procedure also allowed us to examine a) how learning valenced behavioral information influences pain perception (potentially as a function of target race), and further, b) how perceiving a person in pain shapes subsequent impression updating.

### Methods

***Participants.*** 154 participants were recruited from the University of Delaware via SONA, an online recruitment platform used to credit students enrolled in PSYC100 for research participation. No specific inclusionary or exclusionary criteria were used for recruitment, aside from excluding participants in previous pain

perception tasks, including Experiment 1. 24 participants were excluded from analyses due to failure to complete one of the two experimental tasks, thus 130 participants ( $M_{\text{age}} = 18.6$  years,  $SD = 0.804$ ; 95 female, 31 male; 79% White, 10% Black) were used in our final sample.

## **Materials**

**Face stimuli.** For the Impression Formation and Updating tasks, neutral stimuli were the same as used in Experiment 1, though we now selected 48 targets (24 Black, 24 White) in total from the DPD (Mende-Siedlecki et al., 2020). 24 targets appeared in both the Impression Formation and Updating tasks *and* the Pain Perception task, while the other 24 targets appeared in just the former set of tasks.

For the 24 targets appearing in all tasks (i.e., including Pain Perception), we also selected one painful expression from the DPD associated with each of these targets. For the pain perception task, we generated our full set of stimuli by inputting each pair of neutral and painful expressions from each target into Morpheus and generating 11 total morphs per target, resulting in 264 total images used in the Pain Perception task.

**Behavior stimuli.** The behavior stimuli used in the Impression Formation and Updating tasks in Experiment 2 are identical to those used in Experiment 1.

**Procedure.** Participants essentially completed two separate tasks in tandem with each other. As in Experiment 1, participants began each trial with the Impression Formation task. Again, they viewed a neutral Black or White male face paired with either a positive or negative behavioral statement and rated that individual in terms of their trustworthiness, using a scale from -4 to 4 (-4 = very untrustworthy; 4 = very trustworthy).

These individuals were assigned to be in either a “Pain” condition or a “No Pain” condition. In the Pain Condition, after the first trustworthiness rating, participants also completed the Pain Perception task, where they saw 11 face morphs of the same individual they’d just learned about and rated, ranging from a completely neutral expression to a completely painful expression. For each morph, participants were asked to rate how much pain they thought the person was in on a scale from 1 to 7 (1 = definitely not in pain; 7 = definitely in pain); the presentation order of these pain morphs was randomized within blocks, such that painful expressions didn’t increase/decrease linearly. In the “No Pain” condition, participants completed the Impression Formation and Updating tasks, but did not do pain ratings, so as to see how impressions are altered without seeing pain expressions.

In both the “Pain” and “No Pain” conditions, participants then completed the Impression Updating task (i.e., as in Experiment 1), where they viewed the same neutral stimulus from before, but paired with three new morality-related statements and once again provided a trustworthiness rating on a scale from -4 to 4. For targets in both the “Pain” and “No Pain” conditions, behavior in the updating task varied across four experimental conditions: individuals whose behavior changed from negative to positive valence, individuals whose behavior changed from positive to negative valence, individuals who were paired with only positive statements (with no change in valence), and individuals who were paired with only negative statements (with no change in valence). Impression formation, pain rating, and impression updating questions were all self-paced. At the end of the experiment, subjects were asked to complete a demographic questionnaire, although participation wasn’t obligatory.

Participants completed a total of 24 trials each. Blocks were ordered so that presentation of Black and White faces (as well as the directionality of impression updating) would be randomized.

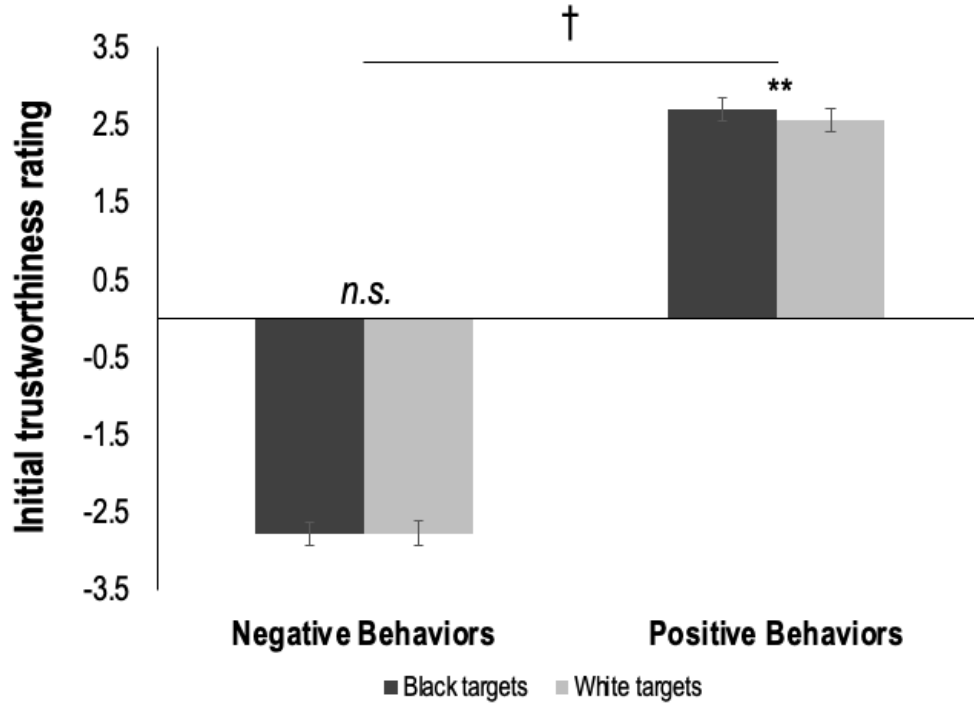
**Analyses.** Our analytic plan comprised three successive steps. First, we conducted a 2 (valence: positive vs. negative)  $\times$  2 (target race: Black vs. White) repeated measures ANOVA to test if the effect of valence on initial impression formation varied by target race. Second, we conducted another 2 (valence: positive vs. negative)  $\times$  2 (target race: Black vs. White) repeated measures ANOVA to test if the effect of initial impression valence on pain perception varied by target race. Third, we conducted a 2 (target race: Black vs. White)  $\times$  2 (initial valence: positive vs. negative)  $\times$  2 (updated valence: positive vs. negative)  $\times$  2 (pain condition: pain vs. no pain)  $\times$  2 (time: first rating vs. second rating) repeated measures ANOVA to see if the effects of updating condition on updating magnitude varied by target race and pain condition.

Subsequently, we tested to see if racial bias in initial impression formation, pain perception, and updating magnitude were correlated with each other.

## Results

**Initial impression task.** As predicted, the main effect of target race was statistically significant ( $F(1,129) = 12.53, p < .001, \eta_p^2 = 0.08$ ); Additionally, the interaction between target race and valence was marginally significant ( $F(1,129) = 5.252, p = .024, \eta_p^2 = 0.039$ ). To further understand these effects, we ran subsequent  $t$ -tests comparing Black and White targets at either level of initial valence. For targets initially paired with negative behaviors, we observed no difference in trustworthiness ratings across race ( $t(129) = 0.982, p = .328, d = 0.086, M_{\text{Black}} = -2.692, SD_{\text{Black}} = 0.921; M_{\text{White}} = -2.737, SD_{\text{White}} = 0.907$ ). Conversely, within targets initially paired

with positive behaviors, Black targets received significantly higher trustworthiness ratings compared to White targets ( $t(129) = 4.457, p < .001, d = 0.391; M_{\text{Black}} = 2.774, SD_{\text{Black}} = 0.85; M_{\text{White}} = 2.59, SD_{\text{White}} = 0.93$ ) (See Figure 3).



**Figure 3. Effects of behavior valence and target race on initial impressions (Experiment 2).** We observed significant effects of race and behavior valence on initial impression formation. Additionally, we found that (like in Experiment 1) Black targets paired with initially positive information were rated higher on trustworthiness than their White counterparts. Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008).  $^{\dagger}p < .10, ^{n.s.}p > .10$

**Updating task.** Despite the predicted five-way interaction including race not being statistically significant, we observed significant four-way interactions between initial valence, updated valence, time, and race ( $F(1,129) = 3.885, p = .051, \eta_p^2 = 0.029$ ), and initial valence, updated valence, time, and pain condition ( $F(1,129) = 7.801, p = .006, \eta_p^2 = 0.057$ ). While the additional effects observed (i.e., a significant main effect of race, a significant interaction between initial valence, updated valence,

and time) are necessarily qualified by these four-way interactions, we focus on breaking each one down separately.

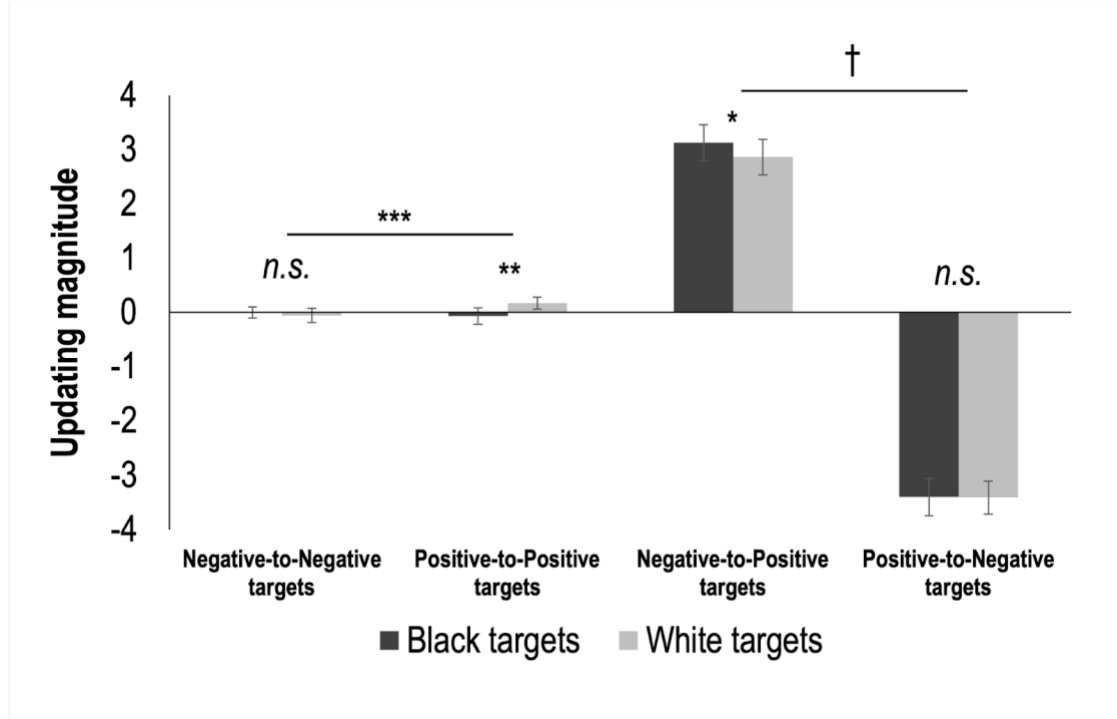
***Four-way interaction between initial valence, updated valence, time, and race.*** To better understand this result, we first tested the three-way interaction between initial valence, time, and race within consistent and inconsistent targets, separately. We observed that this three-way interaction was statistically significant within targets whose behavior was consistent from the first three to the last three behaviors ( $F(1,129) = 6.884, p < .001, \eta_p^2 = 0.051$ ).

Breaking this down further, we split this three-way interaction by valence. Within targets whose behavior was consistently negative, the interaction between race and time was not statistically significant ( $F(1,129) = 0.641, p = .425, \eta_p^2 = 0.005$ ). Participants' trustworthiness ratings stayed similarly negative in response to Black and White targets whose behavior was consistently negative from Time 1 to Time 2 (Black targets:  $t(129) = -0.049, p = .961, d = -0.004, M_{\text{Time1}} = -2.743, SD_{\text{Time1}} = 0.919, M_{\text{Time2}} = -2.740, SD_{\text{Time2}} = 0.986$ ; White targets:  $t(129) = 0.836, p = .405, d = 0.073, M_{\text{Time1}} = -2.697, SD_{\text{Time1}} = 0.990, M_{\text{Time2}} = -2.753, SD_{\text{Time2}} = 1.053$ ).

As for targets whose behavior was consistently positive, we *did* observe a statistically significant interaction between race and time ( $F(1,129) = 9.523, p = .002, \eta_p^2 = 0.069$ ). Participants' trustworthiness ratings stayed stable in response to Black targets whose behavior was consistently positive from Time 1 to Time 2 ( $t(129) = -0.606, p = .546, d = -0.053, M_{\text{Time1}} = 2.773, SD_{\text{Time1}} = 0.884, M_{\text{Time2}} = 2.739, SD_{\text{Time2}} = 0.922$ ). However, their ratings of White targets with the same behavioral profile showed a statistically significant increase over time ( $t(129) = 3.367, p < .001, d = 0.295, M_{\text{Time1}} = 2.544, SD_{\text{Time1}} = 1.024, M_{\text{Time2}} = 2.705, SD_{\text{Time2}} = 1.063$ ).

Within targets whose behavior was *inconsistent* from the first three to last three behaviors, we observed a marginally significant three-way interaction ( $F(1,129) = 3.272, p = .073, \eta_p^2 = 0.025$ ). Breaking this down further, we split this three-way interaction by valence. Within targets whose behavior went from negative to positive, the interaction between race and time was statistically significant ( $F(1,129) = 5.824, p = .017, \eta_p^2 = 0.043$ ). Specifically, changes in trustworthiness ratings from Time 1 to Time 2 were larger for Black targets whose behavior went from negative to positive ( $t(129) = -18.662, p < .001, d = -1.627, M_{\text{Time1}} = -2.642, SD_{\text{Time1}} = 1.068, M_{\text{Time2}} = .485, SD_{\text{Time2}} = 1.668$ ), versus White targets whose behavior went from negative to positive ( $t(129) = -17.295, p < .001, d = -1.517, M_{\text{Time1}} = -2.777, SD_{\text{Time1}} = 1.010, M_{\text{Time2}} = .085, SD_{\text{Time2}} = 1.689$ ).

As for targets whose behavior changed from positive to negative, we did *not* observe a statistically significant interaction between race and time ( $F(1,129) = 0.057, p = .812, \eta_p^2 = < .001$ ). Participants' trustworthiness ratings decreased significantly in response to both Black targets ( $t(129) = 21.825, p < .001, d = 1.914, M_{\text{Time1}} = 2.776, SD_{\text{Time1}} = 0.933, M_{\text{Time2}} = -0.675, SD_{\text{Time2}} = 1.494$ ) and White targets ( $t(129) = 23.064, p < .001, d = 2.023, M_{\text{Time1}} = 2.648, SD_{\text{Time1}} = 0.967, M_{\text{Time2}} = -0.777, SD_{\text{Time2}} = 1.376$ ) whose behavior became changed from negative to positive from Time 1 to Time 2 (See Figure 4).



**Figure 4. Effects of target race and valence on impression updating.** We observed a significant four-way interaction between initial behavior valence, updated behavior valence, race, and time. Additionally, ratings of White targets whose behavior was consistently positive over time saw a larger increase in trustworthiness ratings (as opposed to Black targets). Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008). <sup>†</sup> $p < .10$ , <sup>n.s.</sup> $p > .10$

**Four-way interaction between initial valence, updated valence, time, and pain condition.** Following our approach above we first tested the three-way interaction between initial valence, time, and pain condition within consistent and inconsistent targets, separately. This three-way interaction was statistically significant within targets whose behavior was consistent from the first three to the last three behaviors ( $F(1,129) = 11.495$ ,  $p < .001$ ,  $\eta_p^2 = 0.082$ ).

Breaking this down further, we split this three-way interaction by valence. Within targets whose behavior was consistently negative, we observed a small, but statistically significant two-way interaction between time and pain condition ( $F(1,129) = 4.003$ ,  $p = .048$ ,  $\eta_p^2 = 0.030$ ). Paired t-tests revealed that participants' ratings of consistently negative targets in the "No Pain" condition became marginally worse



from the first rating to the second ( $t(129) = -1.723, p = .087, d = 0.151; M_{\text{Time1}} = -2.723, SD_{\text{Time1}} = 1.002, M_{\text{Time2}} = -2.831, SD_{\text{Time2}} = 0.979$ ), while ratings of consistently negative targets who were observed expressing pain did not change significantly ( $t(129) = 0.877, p = .382, d = 0.077; M_{\text{Time1}} = -2.717, SD_{\text{Time1}} = 0.903, M_{\text{Time2}} = -2.663, SD_{\text{Time2}} = 1.056$ ).

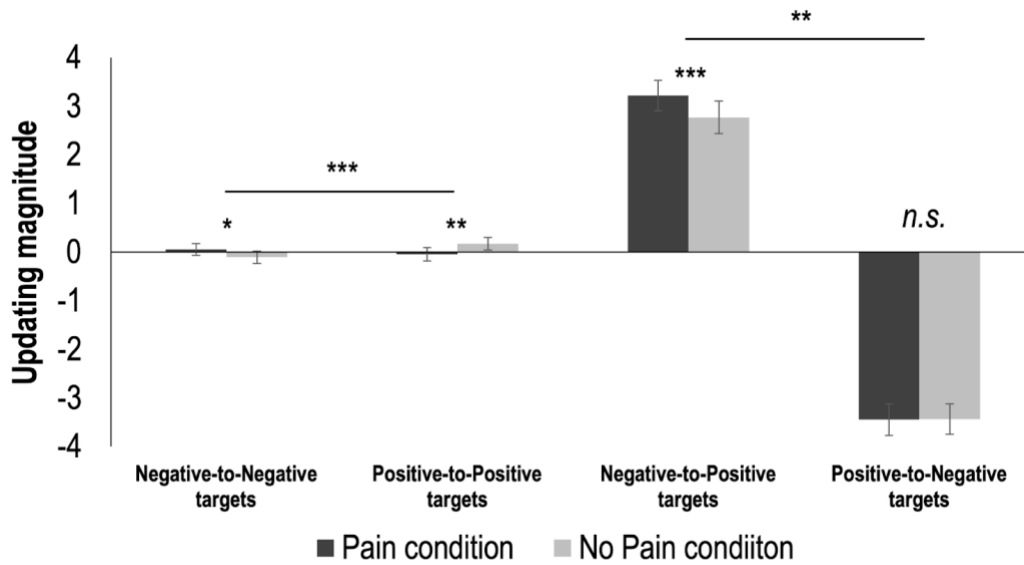
Moreover, within targets whose behavior was consistently positive, we observed another significant interaction between time and pain condition ( $F(1,129) = 7.663, p = .006, \eta_p^2 = 0.056$ ). Paired  $t$ -tests revealed that participants' ratings of consistently positive targets in the "No Pain" condition did not change over time ( $t(129) = -0.773, p = .441, d = -0.068; M_{\text{Time1}} = 2.646, SD_{\text{Time1}} = 0.998, M_{\text{Time2}} = 2.600, SD_{\text{Time2}} = 1.007$ ), while ratings of consistently positive targets who were observed expressing pain became significantly better from the first rating to the second ( $t(129) = 3.099, p = .002, d = 0.272; M_{\text{Time1}} = 2.671, SD_{\text{Time1}} = 0.904, M_{\text{Time2}} = 2.843, SD_{\text{Time2}} = 1.023$ ).

Moving forward to targets whose behavior was *inconsistent* across time, we once again observed a statistically significant three-way interaction between initial valence, time, and pain condition ( $F(1,129) = 9.724, p = .002, \eta_p^2 = 0.070$ ). As we did for the consistent targets, we split this three-way interaction by valence.

First, within targets whose behavior changed from negative to positive, we observed a statistically significant two-way interaction between time and pain condition ( $F(1,129) = 23.963, p < .001, \eta_p^2 = 0.157$ ). Paired  $t$ -tests revealed that participants' ratings of negative to positive targets in the "No Pain" condition increased dramatically from the first rating to the second ( $t(129) = -16.501, p < .001, d = -1.447; M_{\text{Time1}} = -2.754, SD_{\text{Time1}} = 0.994, M_{\text{Time2}} = 0.016, SD_{\text{Time2}} = 1.724$ ). That

being said, this change was even larger within negative-to-positive targets who were also observed expressing pain between the two sets of behaviors ( $t(129) = -20.184, p < .001, d = -1.77; M_{\text{Time1}} = -2.665, SD_{\text{Time1}} = 0.999, M_{\text{Time2}} = 0.554, SD_{\text{Time2}} = 1.724$ ). In other words, impression updates were significantly larger in the negative to positive condition where subjects also viewed targets in pain.

Conversely, for targets whose behavior changed from positive to negative, the interaction between time and pain condition was not statistically significant ( $F(1,129) = 0.009, p = .923, \eta_p^2 < 0.001$ ). Paired  $t$ -tests revealed that participants' ratings of positive-to-negative targets in the "No Pain" condition changed significantly over time ( $t(129) = 22.881, p < .001, d = 2.007; M_{\text{Time1}} = 2.744, SD_{\text{Time1}} = 0.905, M_{\text{Time2}} = -0.689, SD_{\text{Time2}} = 1.466$ ), and further, that the corresponding change in trustworthiness ratings for positive-to-negative targets who were observed expressing pain was comparable in magnitude ( $t(129) = 22.025, p < .001, d = 1.932; M_{\text{Time1}} = 2.680, SD_{\text{Time1}} = 0.963, M_{\text{Time2}} = -0.763, SD_{\text{Time2}} = 1.445$ ) (See Figure 5).

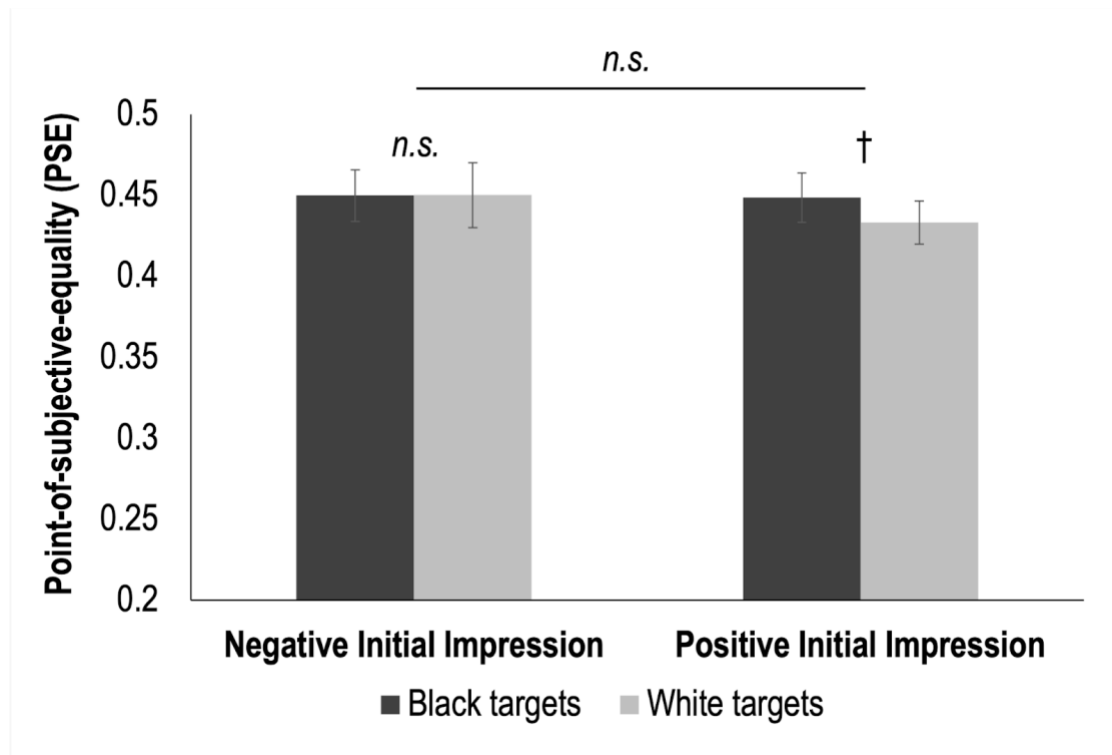


**Figure 5. Effects of behavior valence and pain condition on updating magnitude.** We observed a four-way interaction between initial valence, updated valence, time, and pain condition. Moreover,

seeing pain during scenarios where behavior became increasingly positive led to stronger impression updates. Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008). <sup>†</sup> $p < .10$ , <sup>n.s.</sup> $p > .10$

**Pain perception task.** Neither the main effect of target race ( $F(1,129) = 0.995$ ,  $p = .320$ ,  $\eta_p^2 = 0.008$ ) nor behavior valence ( $F(1,129) = 0.992$ ,  $p = .321$ ,  $\eta_p^2 = 0.008$ ) was statistically significant. While the interaction between race and valence on pain perception thresholds also failed to reach statistical significance ( $F(1,129) = 0.908$ ,  $p = .342$ ,  $\eta_p^2 = 0.007$ ), we broke down the effect of race by valence for exploratory purposes.

We observed that when Black and White faces had been initially paired with negative information, there was no apparent difference in thresholds for perceiving pain on their faces ( $t(139) = 0.023$ ,  $p = .982$ ,  $d = 0.002$ ;  $M_{\text{Black}} = 0.450$ ,  $SD_{\text{Black}} = 0.193$ ,  $M_{\text{White}} = 0.450$ ,  $SD_{\text{White}} = 0.236$ ). However, participants saw pain marginally more readily on White (versus Black) faces that had been initially paired with positive information ( $t(139) = 1.757$ ,  $p = .081$ ,  $d = .154$ ;  $M_{\text{Black}} = 0.448$ ,  $SD_{\text{Black}} = 0.194$ ,  $M_{\text{White}} = 0.433$ ,  $SD_{\text{White}} = 0.179$ ) (See Figure 6).



**Figure 6. Effects of behavior valence and target race on pain perception.** We did not observe significant effects of race or behavior valence on pain perception, nor a significant interaction between those two factors. That said, we note that within the targets initially associated with positive behaviors, we observed a marginally significant simple effect of race, such that participants saw pain somewhat more readily on White (vs. Black) faces. Error bars represent within-subjects corrected 95% confidence intervals (Morey, 2008).  $^{\dagger}p < .10$ ,  $^{n.s.}p > .10$

**Correlational analyses.** We assessed the correlational relationships between bias in pain perception and two separate sources of bias in the impression formation and updating tasks. There was no observed relationship between bias in pain perception (PSE) and bias in initial impressions of targets (all  $r$ s  $< .076$ , all  $p$ s  $> .393$ ). However, participants' tendency to see pain less readily on Black versus White faces was positively and significantly correlated with a tendency to show greater impressions for White (versus Black) targets whose behavior changed from negative to positive ( $r(130) = 0.181$ ,  $p = .039$ ). Moreover, biases in pain perception specifically associated with targets whose behavior was initially negative in valence was positively and marginally correlated with biases in impression updates for targets whose

behavior changed from negative to positive ( $r(130) = 0.153, p = .082$ ). In the updating task, we also observed no significant correlational relationships between tasks (all  $r$ s < 0.128, all  $p$ s > .146).

Taken together, we found that target race and valence each had an effect on impression formation. Specifically, in conditions where targets were paired with initially positive information, Black targets received significantly higher trustworthiness ratings compared to their White counterparts. While race did not appear to impact pain perception in this experiment, we did find an interaction between pain and impression updates. That is, seeing pain on faces appeared to boost trustworthiness ratings in conditions where behaviors were either consistently or increasingly positive (i.e., positive to positive and negative to positive). As in Experiment 1, we observed a notable relationship across tasks that appears to support our overall hypothesis. That is, participants' tendency to see pain more readily on White faces was correlated with stronger impression updates for White targets when their behavior went from negative to positive.

## **Chapter 4**

### **GENERAL DISCUSSION**

In an attempt to test if group-based biases in impression formation and updating are related to other biases in social perception, we conducted two parallel experiments. In our first experiment, participants made initial and updated trustworthiness ratings after learning either positive or negative information about Black and White targets. In addition, they completed a Cross Race Effect (CRE) paradigm to measure face memory for Black and White faces. In alignment with past research, we observed differential sensitivity to faces based on race, such that participants showed worse face memory when making Old/New judgments about Black (versus White) faces. As suggested in past work on the CRE, individuals appear to have overall better memory for White faces, as they correctly recall Old faces more often (“hits”), reject New faces more accurately (“correct rejections”), and have fewer instances of missing or incorrectly labeling faces as Old/New (“misses” and “false alarms”).

In the Impression Formation task, we both replicated past results and observed new (and somewhat unexpected) findings. As seen in previous research, negative information appeared to be more diagnostic and critical for forming impressions, meaning that to individuals, negative information holds more weight than positive information. Adding to this existing literature, we found that this effect was consistent regardless of the target’s race. In the condition where subjects were prompted to make positive initial impressions, however, we observed that trustworthiness ratings were

higher for Black individuals compared to White. In other words, while negative information was regarded as more salient when forming impressions, regardless of race, when learning positive information during impression formation, Black individuals were regarded as even more trustworthy than their White counterparts. This finding contradicts our initial hypothesis that participants would form initially more *negative* impressions of Black targets. It is possible that participants may have simply been attempting to give socially desirable responses, such that they rated Black targets as more trustworthy to make it explicitly clear that the participant does not view Black individuals more negatively than White individuals. Beyond this possibility, it is also the case that we recruited a young, collegiate sample, which is more likely to be socially liberal and endorse more egalitarian views regarding race.

In our Updating task, we also replicated past work by observing clear evidence that participants were updating their impressions: in conditions where behavior changed from one valence to another, impression updates were stronger than conditions where behavior remained consistent. Moreover, we observed evidence that this updating effect was moderated by target valence. When targets' behaviors changed from positive to negative, participants showed bigger changes in their trustworthiness ratings compared to when targets' behaviors changed from negative to positive. Despite our predictions, however, race did not appear to impact how individuals' impressions changed over time. As in the impression formation task, this null effect could have been due to social desirability.

Interestingly, when we compared potential bias across tasks, we observed a relationship between group-based bias in the CRE task and a group-based bias in impression updating specific to negative-to-positive targets. That is, participants who

showed greater sensitivity when making Old/New judgments of White (versus Black) faces in the CRE task also appeared to show relatively stronger impression updates for White (versus Black) targets when behavior went from negative to positive. This suggests that implicit biases may contribute to differential facial perception, which could in turn affect one's evaluation of behavioral information.

In our second experiment, we had participants make initial and updated trustworthiness ratings after learning either positive or negative information about Black and White targets and interlaced a pain perception task to measure participants' sensitivity to painful expressions as a function of target race. Similar to Experiment 1, target race had no impact on negative impression formation, yet when prompted to make a *positive* initial impression, participants rated Black participants as more trustworthy than White targets. Once again, it is possible that this effect stems from either an explicit suppression of prejudicial views, from socially desirable responding, or the liberal-leaning characteristics of our collegiate sample.

Target race not only appeared to affect impression formation, but also updating under certain conditions. When behaviors were consistently positive, participants showed bigger impression updates for White targets. Conversely, when behavior *changed* from negative to positive, updating was stronger for *Black* targets. Thus, in one case where behavior remained unchanged, it appears that racial bias can influence how one perceives another, such that White individuals who behave in a consistently trustworthy manner may be regarded in a more positive light than their Black counterparts, even when their behavior is relatively the same. When behavior *does* change, however, it appears that participants consciously or subconsciously take a target's race into account when making their judgments, and that increasingly positive



behavior from Black individuals is considered more important or salient than the same behavior change in White individuals. That said, it is important to note that these effects of target race were observed when behavior was either consistently positive or when an update was triggered by positive behavior. Within both consistently negative and positive-to-negative targets, where the focus was on more diagnostic (i.e., immoral) behavior, we observed no difference in updating as a function of target race.

Though race did not appear to impact pain perception, seeing pain expressions on people appeared to impact one's impressions. When targets' behavior was consistently negative, seeing them in pain appeared to prevent secondary impressions from becoming even worse than initial ones. Additionally, when targets' behavior was consistently positive, seeing them in pain led to a greater additional boost in impressions than when the targets weren't seen in pain. Thus, in instances where behavior was consistent in valence, seeing pain kept impressions more positive overall than they typically would have been in the same conditions. More critically, when behavior changed from negative to positive, seeing pain on the target's face led to a larger boost in trustworthiness ratings than the "No Pain" control condition. That said, when behavior changed in a more diagnostic direction (i.e., from positive to negative), we saw no change in updating magnitude as a function of pain. Based on these observations, we hypothesize that seeing pain on targets who were initially seen in a negative light might prompt individuation of targets, which may, in turn, promote perception of their pain overall. It is also possible that seeing pain can enhance empathy or satisfy motivations for justice, causing subjects to sympathize with targets to a greater extent and facilitate an improved impression upon learning novel positive information. In other words, if an individual is initially associated with negative

information and then seen in pain, one might feel pity for that person, or feel as though the individual has received a form of “punishment” for their negative actions, causing them to pardon them and update their impression by factoring in both the painful expression and newly learned information.

Most interestingly, we observed a positive correlational relationship between racial bias in pain perception and racial bias within one of our measures of impression updating. More specifically, we found that seeing pain more readily on White faces was correlated with relatively greater impression updates for White (versus Black) targets when their behavior changed from negative to positive. Notably, this correlation directly parallels the finding observed in Experiment 1, where racial bias in the CRE task positively correlated with a bias in this same condition of the updating task. It is notable that in this specific direction of updating, updates are triggered by relatively less diagnostic behaviors and thus, the updating process itself may require more explicit motivation to be engaged (Kim, Park, & Young, 2020).

More generally, the alignment between these two findings provides support for our overarching hypothesis that biases in perception may also be linked to biases in social behavior.

Overall, Experiments 1 and 2 provide support for the theory that there may be relationships between group-based biases in social cognition, memory, and perception. Our results replicate past work on the CRE, demonstrating that our social encoding of others’ faces varies depending on whether they share similar racial identity with us or not. We also observed evidence that this same principle may apply in the context of impression formation and updating, suggesting that we may incorporate race and potential racial prejudice with learned information to form our judgments of people.

Furthermore, the presence of correlational relationships between biases across tasks in each respective experiment also emphasizes our theory that these biases may share a common root, at least to some extent. For example, it's possible that greater perceptual sensitivity to White targets reflects one's ability to better individuate White individuals, which may, in turn, support more well-rounded judgments of White individuals' behavior.

With all of this being said, more work needs to be done to investigate these interactions. It would be especially helpful to compare responses from White subjects to Black subjects to see if these effects are applied only to White targets, or are occurring within same-race targets, aligning with our original suggestions that such biases may derive from general ingroup favoritism and outgroup derogation, rather than beliefs or stereotypes specific to a particular racial group. Additionally, follow-up studies should be done with a more representative sample to hopefully eliminate the social-desirability effect that we believe may have influenced some of our results. Finally, the stimuli used in the experiments could also be replaced with those that better embody people in pain. Seeing as the images we used were all male and showed posed (rather than actually experienced) pain expressions, using stimuli that were more representative of real pain conditions could produce stronger responses.

Individuals' tendency to show racial bias in perception, memory, and impressions all have detrimental consequences, both as independent responses and as a larger, compounded effect. Differential pain perception of Black individuals could explain the systemic lack of medical care and pain management that said individuals experience. Meanwhile a lack of sensitivity to Black faces has been cited as a potential explanation for incorrect eye-witness identifications and subsequent wrongful

convictions. Finally, the influence of racial bias in impression formation and updating could lead to the strengthening of prejudicial stereotypes that impact each of these sectors, in addition to general facilitation of racial group polarization. While our results raise new questions to be addressed, the present work suggests that the everyday microaggressions many people of color experience may be influenced by a myriad of biased behaviors that operate together, rather than in parallel, with one another.

## REFERENCES

- Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Maner, J. K., & Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science*, 17(10), 836–840. <https://doi.org/10.1111/j.1467-9280.2006.01790.x>
- Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The Cross-Category Effect: Mere Social Categorization Is Sufficient to Elicit an Own-Group Bias in Face Recognition. *Psychological Science*, 18(8), 706–712. <https://doi.org/10.1111/j.1467-9280.2007.01964.x>
- Bliss-Moreau, E., Barrett, L., & Wright, C. (2008). Individual Differences in Learning the Affective Value of Others Under Minimal Conditions. *Emotion (Washington, D.C.)*, 8, 479–493. <https://doi.org/10.1037/1528-3542.8.4.479>
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–324. <https://doi.org/10.1037/0033-2909.86.2.307>
- Brewer, M. B. (2010). Social identity complexity and acceptance of diversity. In *The psychology of social and cultural diversity* (pp. 11–33). Wiley Blackwell. <https://doi.org/10.1002/9781444325447.ch2>
- Cassidy, B. S., Krendl, A. C., Stanko, K. A., Rydell, R. J., Young, S. G., & Hugenberg, K. (2017). Configural face processing impacts race disparities in humanization and trust. *Journal of Experimental Social Psychology*, 73, 111–124. <https://doi.org/10.1016/j.jesp.2017.06.018>
- Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us Versus Them: Social Identity Shapes Neural Responses to Intergroup Competition and Harm. *Psychological Science*, 22(3), 306–313. <https://doi.org/10.1177/0956797610397667>
- Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, 114(1), 75–94. <https://doi.org/10.1037/pspi0000110>

Devine, P. G. (19890501). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5.  
<https://doi.org/10.1037/0022-3514.56.1.5>

Dickter, C., & Gyurovski, I. (2012). The effects of expectancy violations on early attention to race in an impression-formation paradigm. *Social Neuroscience*, 7(3), 240–251. <https://doi.org/10.1080/17470919.2011.609906>

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, 33(5), 510–540. <https://doi.org/10.1006/jesp.1997.1331>

Dunbar, R. (2015). Evolutionary Basis of the Social Brain. In *The Oxford handbook of social neuroscience* (First issued as an Oxford University Press paperback, p. 10). Oxford University Press.

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General*, 145(2), 131–146.  
<https://doi.org/10.1037/xge0000132>

Freeman, J. B., Schiller, D., Rule, N. O., & Ambady, N. (2010). The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Human Brain Mapping*, 31(1), 150–159. <https://doi.org/10.1002/hbm.20852>

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In *Unintended thought* (pp. 189–211). The Guilford Press.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience*, 4(8), Article 8. <https://doi.org/10.1038/90565>

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, 69, 669–684. <https://doi.org/10.1037/a0036056>

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology (London, England: 1953)*, 99(Pt 1), 45–56. <https://doi.org/10.1348/000712607X199981>

Hugenberg, K., Young, S., Bernstein, M., & Sacco, D. (2010a). The Categorization-Individuation Model: An Integrative Account of the Other-Race Recognition Deficit. *Psychological Review*, 117, 1168–1187. <https://doi.org/10.1037/a0020463>

- Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010b). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, 117(4), 1168–1187. <https://doi.org/10.1037/a0020463>
- Hughes, B., Camp, N., Gomez, J., & Eberhardt, J. (2019, July 1). *Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception*. <https://doi.org/10.1073/pnas.1822084116>
- Hughes, B., Zaki, J., & Ambady, N. (2016). Motivation alters impression formation and related neural systems | Social Cognitive and Affective Neuroscience | Oxford Academic. *Social Cognitive and Affective Neuroscience*, 12(1), 49–60.
- Kim, M., Park, B., & Young, L. (2020). The Psychology of Motivated versus Rational Impression Updating. *Trends in Cognitive Sciences*, 24(2), 101–111. <https://doi.org/10.1016/j.tics.2019.12.001>
- Kunda, Z., & Thagard, P. (19960101). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, 103(2), 284. <https://doi.org/10.1037/0033-295X.103.2.284>
- Li, T., Cardenas-Iniguez, C., Correll, J., & Cloutier, J. (2016). The impact of motivation on race-based impression formation. *NeuroImage*, 124, 1–7. <https://doi.org/10.1016/j.neuroimage.2015.08.035>
- Meissner, C. A., & Brigham, J. C. (20010307). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3. <https://doi.org/10.1037/1076-8971.7.1.3>
- Mende-Siedlecki, P., Backer, R., & Qu-Lee, J. (n.d.). *Perceptual Contributions to Racial Bias in Pain Recognition*. 27.
- Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic Value Underlies Asymmetric Updating of Impressions in the Morality and Ability Domains. *Journal of Neuroscience*, 33(50), 19406–19415. <https://doi.org/10.1523/JNEUROSCI.2334-13.2013>
- Mende-Siedlecki, P., Goharзад, A., Tuerxuntuoheti, A., Reyes, P. G. M., Lin, J., & Drain, A. (2021). *Assessing the speed, spontaneity, and robustness of racial bias in pain perception*. PsyArXiv. <https://doi.org/10.31234/osf.io/xmdn4>

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharзад, A. (2020). The Delaware Pain Database: A set of painful expressions and corresponding norming data. *Pain Reports*, 5(6), e853. <https://doi.org/10.1097/PR9.0000000000000853>

Mende-Siedlecki, P., & Todorov, A. (2016). Neural dissociations between meaningful and mere inconsistency in impression updating. *Social Cognitive and Affective Neuroscience*, 11(9), 1489–1500. <https://doi.org/10.1093/scan/nsw058>

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience & Biobehavioral Reviews*, 37(8), 1530–1536. <https://doi.org/10.1016/j.neubiorev.2013.06.002>

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior*, 35(3), 169–175. <https://doi.org/10.1016/j.evolhumbehav.2013.12.006>

Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology*, 32(5), 460–473. <https://doi.org/10.1037/a0029203>

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, 108(19), 7710–7715. <https://doi.org/10.1073/pnas.1014345108>

Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1589), 744–753. <https://doi.org/10.1098/rstb.2011.0300>

Stern, S. E., Mullennix, J. W., Corneille, O., & Huart, J. (2007). Distortions in the memory of the pitch of speech. *Experimental Psychology*, 54(2), 148–160. <https://doi.org/10.1027/1618-3169.54.2.148>

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149–178. <https://doi.org/10.1002/ejsp.2420010202>

Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & Psychophysics*, 69(4), 619–627. <https://doi.org/10.3758/BF03193919>



Todorov, A., & Olson, I. R. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. *Social Cognitive and Affective Neuroscience*, 3(3), 195–203. <https://doi.org/10.1093/scan/nsn013>

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83(5), 1051–1065. <https://doi.org/10.1037/0022-3514.83.5.1051>

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39(6), 549–562. [https://doi.org/10.1016/S0022-1031\(03\)00059-3](https://doi.org/10.1016/S0022-1031(03)00059-3)

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science*, 19(11), 1131–1139. <https://doi.org/10.1111/j.1467-9280.2008.02214.x>

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The cross-race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review*, 7(1), 83–113. <https://doi.org/10.1111/j.1751-2409.2012.01044.x>

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, 47(2), 237–252. <https://doi.org/10.1037/0022-3514.47.2.237>

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, 29(26), 8525–8529. Scopus. <https://doi.org/10.1523/JNEUROSCI.2418-09.2009>

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, 16(2), 116–142. <https://doi.org/10.1177/1088868311418987>

## APPENDIX



**Institutional Review Board**  
210H Hulihan Hall  
Newark, DE 19716  
Phone: 302-831-2137  
Fax: 302-831-2828

DATE: November 17, 2022

TO: Azaadeh Goharзад  
FROM: University of Delaware IRB

STUDY TITLE: [1974010-1] Integrating Face and Behavior Information in Social Judgments  
SUBMISSION TYPE: New Project

ACTION: APPROVED  
EFFECTIVE DATE: November 17, 2022  
NEXT REPORT DUE: November 16, 2023

REVIEW TYPE: Expedited Review  
REVIEW CATEGORY: Expedited review category # (7)

Thank you for your New Project submission to the University of Delaware Institutional Review Board (UD IRB). The UD IRB has reviewed and APPROVED the proposed research and submitted documents via Expedited Review in compliance with the pertinent federal regulations.

As the Principal Investigator for this study, you are responsible for, and agree that:

- All research must be conducted in accordance with the protocol and all other study forms as approved in this submission. Any revisions to the approved study procedures or documents must be reviewed and approved by the IRB prior to their implementation. Please use the UD amendment form to request the review of any changes to approved study procedures or documents.
- Informed consent is a process that must allow prospective participants sufficient opportunity to discuss and consider whether to participate. IRB-approved and stamped consent documents must be used when enrolling participants and a written copy shall be given to the person signing the informed consent form.
- Unanticipated problems, serious adverse events involving risk to participants, and all non-compliance issues must be reported to this office in a timely fashion according with the UD requirements for reportable events. All sponsor reporting requirements must also be followed.

The UD IRB REQUIRES the submission of a PROGRESS REPORT DUE ON November 16, 2023. A continuing review/progress report form must be submitted to the UD IRB at least 45 days prior to the due date to allow for the review of that report.

If you have any questions, please contact the UD IRB Office at (302) 831-2137 or via email at [hsrb-research@udel.edu](mailto:hsrb-research@udel.edu). Please include the study title and reference number in all correspondence with this office.

**INSTITUTIONAL REVIEW BOARD**

- 1 -

Generated on IRBNet