SOCIAL NEUROSCIENCE OF INTERGROUP DYNAMICS

by

Samuel A. Venezia

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychological & Brain Sciences

Spring 2025

SOCIAL NEUROSCIENCE OF

INTERGROUP DYNAMICS

by

Samuel A. Venezia

Approved: _____

Robert West, Ph.D.
Chair of the Department of Psychological & Brain Sciences

Approved: _____

Caleb Everett, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Jennifer T. Kubota, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Jasmin Cloutier, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Naomi Samimi-Sadeh, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Pete Simi, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to thank both of my advisors, Dr. Jennifer Kubota and Dr. Jasmin Cloutier for their feedback and support during this eight year long journey to my PhD. Thank you for always being available for advice, either personal or professional, and for always helping me to be a better scientist. Thank you to Dr. Naomi Sadeh-Samimi and Dr. Peter Simi for your time and thoughtful critiques to this dissertation. Thank you again, Dr. Jim Davis and Dr. Bradley Mattan, for your guidance and support in the early days of my academic career. I would not be in a graduate program today if it were not for you two. I also want to thank Frank Novakowski, for where I am today. Frank, you were the first person to believe that I had potential as a scientist. I would not be the scientist I am today without your initial mentorship. I also want to thank the IFSN Lab graduate students who graduated before me, Dr. Tzipporah Dang, Dr. Grace Handley, Dr. Denise Barth, Dr. Anjana Lakshmi, and Dr. Kevin Wei, thank you all for your mentorship, guidance and friendship. Thank you to Dr. Eric Splan for the mentorship and guidance on the network/ multivariate collaborations.

I also wish to acknowledge the rest of the IFSN Lab for the time I have spent with you all over the last eight years. Thank you, Richa Gautam, for your insight, and support during our shared coursework as well as the ARL grant study over the past five years. Thank you Inshil Addison for your hard work and perseverance on the ARL grant; your stepping up with data collection efforts was integral for the work the grant covered, and I am excited to see what research comes from these efforts. I also wish to acknowledge Melda Sandikci, Marissa Spears, Samwell Cleary, Maggie Welte

and Giuditta Scalco. I wish you all the best as you continue on your PhD career, and I am excited to see where your research goes. I would also like to very graciously thank Andrea Wilhelm, Roxanna Moosavy, Francesca Tero, Andréa Reynolds, Nick Perullo, and Alyssa Diaferia for their continued support with data collection for this dissertation. Andrea, Andréa, Nick, and Roxanna, in particular, I cannot express how integral your help has been toward the completion of this dissertation; I am forever grateful for your help. Andrea, it has also been an honor to be your mentor during my time in the IFSN lab. I know you will be a truly great academic, and I could not be more grateful to be your friend outside of the lab. Audie, it has also been a privilege to become close friends with you and Andrea and cannot express how grateful I am to know that I've made lifelong friendships with both you and Andrea during my time in Delaware.

To both of my parents, Tony and Becky, and my sister Naomi, thank you for your never-ending love and support, this has kept me going particularly during my last two years of this journey toward my PhD. To my cat Tamar, thank you for being the loveable goofball you are that has provided laughter and companionship when things were their toughest. Last but certainly not least, thank you to my wife, favorite travel buddy, and most of all, best friend, Ginny. Thank you for your love, support, encouragement, proofreading, encouraging me to persevere even when things were at their hardest, and always being able to keep me grounded.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Tajfel and Turner's Social Identity Theory characterizes how individuals form ingroups and outgroups and how affinity to one's ingroup and degradation of outgroups contribute to the formation of intergroup biases. Research, including neuroimaging studies, examining intergroup dynamics have primarily used face stimuli, potentially overlooking meaningful information gleaned from social interactions. As such, this dissertation uses fMRI to examine the neural correlates of intergroup processing of complex depictions of affiliative and aversive social behaviors. Study 1A curated stimuli depicting affiliative social interactions (i.e., perceived Black, Interracial, and White couples). Study 1B curated stimuli depicting aversive social interactions (i.e., depictions of violence toward Black or White victims) and depictions of mundane social interactions. Study 2 focused on brain regions supporting affiliative intergroup evaluations and salience processing among a sample of self-identified White Americans (n = 49). Results found greater activity in regions supporting positive evaluations (i.e., OFC, NACC) and salience (amygdala, AI) toward depictions of Interracial couples relative to depictions of perceived Black or perceived White couples. This suggests that while perceivers may view Interracial couples as more salient compared to White or Black couples, they also evaluated them more positively. Study 3 examined brain regions supporting mentalizing (i.e., STS, TPJ) and aversion (AI) while processing aversive intergroup behaviors using the same sample as Study 2. Results found greater activity in the STS and AI toward the mundane scenes relative to depictions of White victims. Between violent conditions,

results found more activity in the AI, suggesting more aversive responses to depictions of an outgroup member being harmed. There was also greater activity in the TPJ to depictions of Black victims relative to White victims, suggesting that perceivers engaged in more effortful mentalizing toward Black relative to White victims. Results from Studies 2 and 3 are further discussed relative to their implications for intergroup impression formation and intergroup bias for existing work rooted in Social Identity Theory.

Chapter 1

INTRODUCTION

Social Identity Theory has characterized intergroup dynamics as those supporting the categorization of others into ingroups (individuals/groups who are similar to the self) and outgroups (individuals/groups who are different from the self). Within Social Identity Theory, such spontaneous categorizations of others into ingroups and outgroups may also make salient associations or explicitly held beliefs one may have about ingroups and outgroups. Moreover, self-categorization into ingroups is thought to provide individuals with both self-esteem as well as a distinct identity (Tajfel, 1982; Tajfel et al., 1971; Tajfel & Turner, 1979). Such positive self-associations reinforce positive appraisals of ingroup members, often referred to as ingroup love (Allport, 1954; Brewer, 1999). Affinity for one's ingroup may not always be directly attributable to the formation of intergroup biases, as it has been thought that one may have a sense of ingroup love without necessarily having animus toward outgroups (Allport, 1954; Brewer, 1999, 2017; Halevy et al., 2012). The presence of outgroup hate is generally not considered to spontaneously arise from the presence of ingroup love, particularly if group norms discourage animus toward outgroups (e.g., Crandall et al., 2002; Iacoviello & Spears, 2018). However, in the presence of perceived threats or conflict between one's ingroup and other outgroups (e.g.,

perceived scarcity of resources, real or symbolic threats to one's ingroup's wellbeing) strong ingroup love may has been thought to contribute to the formation of bias toward outgroups (De Dreu, 2010; Halevy et al., 2008, 2012). Relatedly, perceptions of other groups being lower in a social hierarchy, perceptions of threat or conflict from outgroups, or differences in outgroup social structures may be related to greater degradation of members of outgroups or outgroup hate, that may also contribute to the formation of intergroup bias (M. T. Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015).

Since the cognitive revolution, social psychology and social neuroscience have meticulously examined the intergroup dynamics that are shaped by ingroup love and outgroup hate. More specifically, within either field, there exists a large body of work that examines how ingroup love and outgroup hate may shape evaluations of other groups (e.g., Cikara et al., 2011; Mattan et al., 2019; Mattan, Kubota, Dang, et al., 2018; Zhou et al., 2020), associations held about members of other groups (e.g., Dovidio et al., 2002; Dunsmoor et al., 2016; Greenwald et al., 1998), and how perceivers interact with members of other groups (e.g., Hein et al., 2010; Kubota et al., 2013, 2023; Stanley et al., 2012).

While efforts to further explore intergroup dynamics have been incredibly successful, one inherent shortcoming of investigations where group identities are gleaned from perceived race is that the stimuli utilized consist primarily of faces (e.g., Cloutier et al., 2017; Handley et al., 2022; Kubota & Ito, 2007; Mattan et al., 2019; Mattan, Kubota, Dang, et al., 2018; Wilson et al., 2017), rather than stimuli that depict

individuals engaging in social behaviors. While face stimuli offer perceptual information that makes the inference of group identity (e.g., perceived race) possible with little effort, face stimuli do not offer information as to how individuals may process interactions between members of different groups, particularly if the nature of the interactions are affiliative or aversive. Consideration for more complex depictions of social interactions between members of different groups may help further contextualize how perceivers interpret behaviors between people of different group identities. In turn, this may highlight meaningful nuance as to how evaluations and associations of ingroup and outgroup members may change based on the behaviors displayed. The additional context provided by more depictions of complex social behaviors may be able to further elucidate how intergroup processing may relate to real-world instances of intergroup biases and discrimination (e.g., Burke et al., 2017; Derous et al., 2015; Simi et al., 2017; Simi & Futrell, 2010).

To this end, the present dissertation seeks to address this gap in the literature and examines the neural correlates of mechanisms supporting intergroup processing when perceivers view and form impressions of stimuli depicting more complex social interactions between ingroup and outgroup members.

Why Social Neuroscience?

Social neuroscience is a discipline that examines the link between social behaviors and the brain and allows researchers to link neural mechanisms to psychological processes. Social neuroscience allows researchers to consider a wide

variety of factors that may explain intergroup dynamics (e.g., affect, cognition, motivation). Because of its ability to consider a wide variety of factors, social neuroscience has been heavily utilized to examine how ingroup love and outgroup hate may influence how people evaluate and infer the thoughts and intentions of members of other groups and how perceivers may regulate their attitudes and behaviors toward members of different groups (see Amodio & Cikara, 2021; Kubota et al., 2012 for reviews).

In addition to its ability to consider how a wide range of psychological factors may shape intergroup dynamics, social neuroscience offers additional advantages over behavioral methods. For example, social neuroscience is able to circumvent the fact that people may often not be able to accurately introspect or choose not to do so based on factors such as social desirability (Amodio, 2010; Amodio & Cikara, 2021; Bartholow & Dickter, 2007; Berkman et al., 2014; Lieberman, 2007).

Another relative advantage of social neuroscience is that it allows researchers to consider multiple psychological processes within a single paradigm. For example, if a sample of perceivers categorized faces of ingroup members faster than outgroup members based on reaction time data, it may suggest a general pro-ingroup bias. However, in the absence of other behavioral measures (e.g., explicit ratings of the faces), it may be difficult to infer what was driving the more efficient processing of ingroup members. Social neuroscience, however, would allow researchers to examine differences in brain activity during the processing of others, which could elucidate

additional mechanisms that may explain observed results beyond what behavioral data may be able to.

Social neuroscience's benefits of being able to access psychological mechanisms that are not accessible to perceivers, as well as its ability to track multiple processes simultaneously, is of particular relevance to this dissertation. For example, when perceivers process members of ingroups and outgroups, they may evaluate others, pay attention to them, or try to infer their thoughts and intentions more or less depending on their group identity (Amodio & Cikara, 2021; Kubota et al., 2012; Mattan, Wei, et al., 2018). Consideration for depictions of social behaviors may further shape how perceivers process interactions between ingroup members and outgroup members. On one hand, viewing affiliative social interactions may be related to differential engagement of brain networks supporting positive evaluations and ones that support attention and salience processing, depending on whether people in the images are perceived as ingroup members or outgroup members. Conversely, viewing aversive social interaction may be related to differential engagement of brain networks supporting salience or negative evaluations as well as those supporting the inference of mental states from others depending on whether the individuals in the scenes are perceived as ingroup or outgroup members.

Social neuroscientific methods provide strong advantages for research examining intergroup dynamics. Therefore, the present dissertation examines brain regions that support positive social evaluations/reward, salience/aversion, inference of mental states from others (i.e., mentalizing), and in an exploratory manner, regions

that support the regulation of intergroup bias when a sample of White perceivers process stimuli depicting affiliative or aversive behaviors.

In the following section, we provide an initial review of background literature surrounding the social neuroscience of intergroup dynamics with respect to brain regions supporting reward, salience/aversion, mentalizing, and the regulation of intergroup bias. After the initial discussion of relevant brain regions, we present Study 1A, where we curated an equated set of images that depict happy, smiling couples comprised of men and women of different race compositions. Next, in Study 1B, we detail how we curated and equated a set of images depicting aversive social interactions (i.e., depictions of violence toward either a Black or White victim by a Whtie aggressor). Then we discuss Study 2, wherein a sample of White perceivers viewed and formed impressions of the images of couples that were curated in Study 1A while undergoing fMRI. Following Study 2, we detail Study 3, where the same sample of participants from Study 2 formed impressions of the aversive social interaction while undergoing fMRI. Finally, we close with a general discussion of the two neuroimaging studies, their contributions to theory, future directions, and study limitations.

The Neural Correlates of Intergroup Processing

Reward in Intergroup Bias

Intergroup evaluations are the valenced associations held about groups. These evaluative associations (positive, negative, ambivalent) can be about ingroups, outgroups, or both. Positive evaluations are typically held about one's ingroup (Brewer, 1999), and these positive associations with the ingroup may amplify one's self-worth (Tajfel & Turner, 1986). Within existing social psychological frameworks, this positive appraisal, or ingroup love, is generally thought to uniquely lead to ingroup favoritism (or bias; Brewer, 1979; Tajfel & Turner, 1986)and can influence intergroup discrimination without the presence of outgroup animus (Brewer, 2017). Previous neuroimaging research supports a potential ingroup preference interpretation because viewing, interacting with, or getting rewards from ingroup members (relative to outgroup members) is associated with preferential activity in regions involved in reward processing (Hackel et al., 2017; Telzer et al., 2015). More specifically, intergroup neuroimaging research has found that viewing or responding to ingroup members (relative to outgroup members) is related to activity in the reward network, including the nucleus accumbens (NACC) and the orbital frontal cortex (OFC). In light of the neuroscientific evidence showing enhanced positive evaluations of ingroup relative to outgroup members in regions within the reward network, these regions may likely play a role in reinforcing and strengthening explicit favoritism.

Nucleus Accumbens

Within existing neuroscientific work examining the underpinnings of reward processing, the NACC is implicated in the anticipation and processing of stimuli that may evoke a pleasure response, such as appetizing foods (Beaver et al., 2006; Mahler et al., 2007; Peciña & Berridge, 2005) or drugs (Childress et al., 2008; Peciña et al., 2006). Additionally, the NACC has been broadly implicated in processing subjective value (Heekeren et al., 2007; Ly et al., 2011; Zink et al., 2008). Crucially, this function of the NACC assists with social learning by facilitating the maximization of future rewards and minimizing losses (Delgado et al., 2005; Fareri & Delgado, 2014). Within the social domain, it has been proposed that the NACC may play a role in ingroup favoritism through instrumental learning. Specifically, the NACC has been observed to activate preferentially to ingroup members during social evaluations (Cikara et al., 2011; Hackel et al., 2017, 2020). Enhanced ingroup positive evaluations have also been observed as a function of viewing depictions of perceived racial ingroup members, with preferential NACC activity being elicited to ingroup members when perceivers viewed cross-race social interactions (Stanley et al., 2012; Wang et al., 2019).

Although work on the neuroscience of intergroup reward generally finds patterns of activity in line with ingroup "love," there is evidence that supports reward responses to outgroup "hate." For example, research on intergroup schadenfreude (i.e.,

one taking pleasure at another's misfortune) has revealed activity in the NACC when people witness a disliked outgroup experience a negative outcome (Cikara et al., 2011; Cikara & Fiske, 2012, 2013; L. T. Harris et al., 2010; Singer et al., 2006; Takahashi et al., 2009). Researchers have also observed that the reward responses in NACC are significantly correlated with participants' desires to commit acts of violence toward members of disliked outgroups (Cikara et al., 2011) and a general desire to take revenge on outgroup members for committing actions that harmed one's ingroup (Singer et al., 2006).

OFC

The OFC has been generally associated with the generation of affective meaning (Rolls & Grabenhorst, 2008; Roy et al., 2012), including value processing (Berridge & Kringelbach, 2008; Kringelbach & Berridge, 2009) as well as social evaluations (Mattan et al., 2017; Mende-Siedlecki et al., 2013). Relatedly, the OFC has been thought to support perceptions of trust when evaluating others (Cassidy & Krendl, 2016; Delgado et al., 2005; Willis et al., 2010). The OFC has also been thought to reflect more positive evaluations of ingroup members relative to outgroup members (Amodio, 2014; Gilbert et al., 2012; Van Bavel et al., 2008). Beyond group identity derived from race, the OFC, and more specifically, the ventromedial prefrontal cortex (VMPFC), a sub-region within the OFC has been found to be associated with social evaluations based on person-knowledge (Cloutier et al., 2012; Cloutier & Gyurovski, 2014; Dang et al., 2019) such that greater engagement of the

VMPFC is associated with positive social evaluations of conspecifics for which people possess knowledge. For example, the VMPFC has been observed to have preferential activity toward people who are relatively high in perceived moral status (Cloutier et al., 2012; Cloutier & Gyurovski, 2014) and positively evaluated familiar people (Dang et al., 2019). Despite evidence implicating the VMPFC in knowledge-based social evaluations, there is limited evidence of its differential involvement based on intergroup identities (c.f., Morrison et al., 2012). Indeed, the VMPFC appears not to be consistently preferentially responsive to ingroup evaluations (Mattan, Kubota, Dang, et al., 2018). However, Firat and colleagues (2017) did find preferential VMPFC activity toward perceived White relative to perceived Black people. Furthermore, it is noteworthy that most brain-imaging investigations of intergroup social evaluation were based on novel others rather than people for which the perceiver possesses person-knowledge.

Accordingly, the limited evidence on the VMPFC's role in intergroup social evaluations does not necessarily mean it should be ignored in future investigations on intergroup bias. Because of the OFC's role in generating affective meaning (Roy et al., 2012), it could also support the generation of intergroup biases. Based on the evidence supporting the role of the VMPFC in generating positive social evaluations based on person knowledge (Cloutier et al., 2012; Cloutier & Gyurovski, 2014; Dang et al., 2019), it is possible that evaluations based on intergroup knowledge may reveal preferential VMPFC activity to ingroup members. In turn, the salience of negative associations toward outgroup relative to ingroup members may be associated with

preferential activity indexing negative evaluations (Tamir & Hughes, 2018). However, due to the limited evidence that would link the VMPFC to race-based social evaluations, this dissertation will primarily focus on the OFC as it has shown a more consistent implication in intergroup social evaluations from race information relative to the VMPFC.

Summary: Intergroup Biases and Reward

Taken together, existing literature concerning the role of brain regions associated with reward and social evaluation suggests that these processes are associated with intergroup bias. More specifically, regions such as the NACC and OFC may support the generation of pro-ingroup biases. They may also support anti-outgroup biases through generating affective meaning based on social category information.

The Role of Salience Detection in Intergroup Bias

While positive intergroup evaluations can serve as a mechanism underlying pro-ingroup bias, negative evaluations are also frequently associated with intergroup biases. Positive intergroup evaluations may impact intergroup discrimination independently of outgroup animus (Brewer, 2017). However, it is difficult to ignore the role of negative evaluations of outgroups within the context of intergroup dynamics. Neuroscientific investigations reveal examples of these anti-outgroup biases. For example, the amygdala, a brain region that has been associated with

saliency detection, has shown preferential responses when viewing outgroup members relative to when viewing ingroup members (Hart et al., 2000; Krill & Platek, 2009; Lieberman et al., 2005; Richeson et al., 2008; Stanley et al., 2012). Additionally, the anterior insula (AI), a brain region associated with negative evaluative processing, has shown preferential activity when interacting with members of outgroups (Merritt et al., 2021; Shkurko, 2013). Based on these findings, it is also possible that these regions may, at least in part, shape outgroup degradation.

Amygdala

The amygdala has been associated with fear responses across humans and animals (Olsson et al., 2007; Phelps, 2006; Phelps & LeDoux, 2005). In the person perception literature, increased activity in the amygdala has generally been associated with rapid perceptions of salient social information (Adolphs, 2009; Stanley et al., 2012). Early literature concerning race perception found greater amygdala activity toward perceived Black relative to perceived White faces among White perceivers during rapid face presentation (Cunningham et al., 2004). Studies investigating race perception characterized enhanced amygdala activity toward outgroup members as representing direct negative social evaluations (Phelps et al., 2000) or as possibly underlying perceptions of threat (Lieberman et al., 2005). However, the amygdala's relation to intergroup bias is likely less simple than being inflexibly implicated in threat or fear perceptions toward outgroup members. For example, several neuroscientific investigations on intergroup bias have not observed significant

differences in activity as a function of whether a face was perceived to be an ingroup or outgroup member (Beer et al., 2008; Dang et al., 2022; Handley et al., 2023; Mattan, Kubota, Dang, et al., 2018; Phelps et al., 2000).

Moreover, it is not completely clear as to whether the amygdala was involved with salience processing due to perceived negative evaluation or novelty (Lieberman et al., 2005). Some investigations suggest that the amygdala may reflect salience toward outgroup members as a result of novelty (Van Bavel et al., 2008) and that greater expertise with outgroup members (e.g., intergroup contact) can reduce amygdala activity toward outgroup members (Cloutier, Li, et al., 2014). Taken together, this would suggest that the amygdala does not directly underly negative evaluations of outgroups; rather, it seems more likely that the amygdala may be associated with general salience processing and support the learning and generalization of negative evaluations of outgroups rather than intergroup biases themselves.

Molapour and colleagues (2015) found that participants who underwent fear conditioning with perceived White and perceived Black faces showed enhanced amygdala activity to outgroup faces that they were learned to associate with negative outcomes (i.e., pain from an electrical shock). This investigation also found that among their sample of White participants, the increased amygdala activity predicted pro-white explicit bias such that participants with increased amygdala activity subsequently interacted with perceived White people but deliberately excluded perceived Black people. The amygdala has also been shown to reflect culturally

learned intergroup associations (Chekroud et al., 2014). This can be seen in studies that found that greater implicit negative associations towards perceived Black compared to perceived White faces were associated with enhanced amygdala activity toward perceived Black faces among White perceivers (Cunningham et al., 2004; Krill & Platek, 2009; Phelps et al., 2000). However, the amygdala may also reflect learned social threats related to appearing prejudiced in addition to intergroup associations (Mattan, Kubota, Dang, et al., 2018). For example, Mattan and colleagues (2018) investigated how people's external motivation to respond without prejudice (EMS; Plant & Devine, 1998) shaped their response to race and social status information. People with high levels of EMS tend to be motivated by external pressures to maintain egalitarian attitudes rather than an intrinsic motivation to not act in a prejudiced manner (Butz & Plant, 2009). While Mattan and colleagues did not find preferential amygdala activity as a function of perceived race, they did find that their participants who were high in EMS showed greater amygdala activity as a function of the other readily available social category information (i.e., social status). This pattern of results was attributed to the presence of social status information being present in addition to perceived race. In the case of their findings, they believed that in the presence of multiple social categories, high-EMS participants showed amygdala activity that may have been related to greater salience of low-SES faces rather than as a function of perceived race.

Anterior Insula

Beyond the amygdala, the AI is a brain region that could provide unique insights into how aversion towards outgroup members may rise from intergroup biases. Implicated in a wide range of functions, the AI serves as a hub for saliency detection regions (Menon & Uddin, 2010). Within an intergroup context, the AI has been implicated in salience detection of ingroup coalition building (Lau et al., 2020). Specifically, Lau et al., found that the AI was particularly implicated in salience monitoring and detection during the categorization of novel people. Their results suggested that activity in the AI was a robust predictor of categorizing novel people as allies or not (i.e., ingroup or outgroup members). Existing studies have also found that the AI has also been associated with processing perceptually novel faces. In particular, the AI has been thought to be implicated in processing novel faces due to the increased saliency of novel people (Cloutier et al., 2017; Cloutier, Li, et al., 2014). The AI has also been shown to be sensitive to social salience beyond perceptual information and has been implicated in social categorizations based on salient person-knowledge (e.g., political orientation; Cikara et al., 2017). Taken together, these results suggest that the AI is responsible for coordinating saliency detection and integrating social information into other brain regions associated with saliency processing.

In addition to its role in saliency detection, the AI has been implicated in processing aversive stimuli and is frequently tied to perceived disgust (Borg et al., 2008; Moll et al., 2005; Wicker et al., 2003). For example, the AI has been previously tied to disgust or aversive responses when viewing pathogenic stimuli (Calder et al.,

2007; Heining et al., 2003); it has also shown sensitivity towards perceived moral violations that others (Borg et al., 2008; Ying et al., 2018) or even oneself (Fourie et al., 2017) might commit in the future (Chapman & Anderson, 2013). The AI may also show preferential activity to norm violations among people who have intergroup biases. For example, Skinner & Hudac (2017) examined how White perceivers processed images of couples that were either racially homogenous (i.e., both partners were perceived to be White) or where one partner was perceived to be White and the other perceived to be Black. Their results suggest preferential activity within the AI to Interracial couples rather than same-race couples, regardless of whether people had explicitly biased views on Interracial dating, as indicated by explicit ratings of their approval of Interracial dating and marriage (Skinner & Hudac, 2017). As such, the AI may be particularly sensitive to depictions of outgroup members that are found to be salient (i.e., aversive).

Summary: Salience, Aversion and Intergroup Bias

Taken as a whole, existing literature discussing regions implicated in salience detection in an intergroup context has generally suggested greater activation of the amygdala and AI to depictions of outgroup members relative to ingroup members. These findings have been thought to be related to greater perceived salience of outgroup members relative to ingroup members. Furthermore, literature examining intergroup aversion has found the AI tends to activate more toward depictions of

outgroup members relative to ingroup members. These results suggest that both the amygdala and AI may support the generation and sustainment of anti-outgroup biases.

Intergroup Bias and Mentalizing

TPJ and DMPFC

Neuroscientific work has suggested that both the temporoparietal junction (TPJ; Koster-Hale et al., 2017; Saxe & Kanwisher, 2003; Schlaffke et al., 2015; Spreng et al., 2009; Young et al., 2011; Young & Saxe, 2009) and the dorsomedial prefrontal cortex (DMPFC; Handley et al., 2022; Mitchell et al., 2005) are associated with the process of inferring mental states from others. Specifically, when making inferences about others' mental states from complex social situations, people demonstrate increased activity in the TPJ relative to when tasked with physically describing others or making judgments about non-social situations (Saxe & Kanwisher, 2003). A recent study found enhanced TPJ and DMPFC activity when participants were tasked with making social relative to non-social judgments of faces that differed in race (i.e., how interested or symmetrical each face appeared to be; Handley et al., 2022). The reduction in DMPFC activity has also been associated with reduced mentalizing towards outgroup members (Bruneau et al., 2018; L. T. Harris & Fiske, 2006). For example, people may exhibit a reduction in DMPFC activity when tasked with evaluating groups that are viewed negatively by the perceiver (e.g., images of unhoused people). Whereas both the TPJ and DMPFC are believed to positively

index greater mentalizing toward others, the TPJ has also been suggested to be implicated with taking another's perspective. Martin and colleagues (2020) found that direct stimulation of the TPJ using transcranial direct current stimulation predicted enhanced perspective-taking with an avatar, whereas stimulation to the DMPFC did not. This study implies that while either region may be implicated in mentalizing, each region's role in mentalizing may be separable. Indeed, the TPJ may play a larger role in supporting perspective-taking than other forms of mentalizing (e.g., mental state inference).

The TPJ has also been found to be recruited when processing person-knowledge about others that violates expectations (Cloutier et al., 2011). For example, a study found that participants' greater TPJ activity was associated with diminished punishments for ingroup members who violated social norms (Baumgartner et al., 2012). The findings of this study indicate that people may be more likely to mentalize in response to ingroup relative to outgroup members when motivated to understand their behavior. Similarly, the TPJ may moderate intergroup preferences in punishment. In one investigation (Baumgartner et al., 2014), participants observed a dyad (i.e., one person was an ingroup member, the other an outgroup member) playing a prisoner's dilemma task after having undergone actual or sham transcranial magnetic stimulation (TMS) on their TPJ. In the task, participants observed two partners complete a decision-making task together and were able to punish either partner if they felt they acted unfairly toward the other partner. Results found that people who received TMS on their right TPJ did not show preferential punishment toward outgroup members,

whereas those who received sham TMS did punish outgroup members more. The authors frame these findings as greater disruption of the TPJ, which was related to diminished consideration for the mental states of ingroup members, which they suggested may lead to more leniency in punishments toward ingroup members. Therefore, these results suggest that when the TPJ shows preferential activity to ingroup members, it may be related to greater mentalizing toward ingroup members relative to outgroup members.

While much of the literature reviewed in this section has focused on the TPJ's sensitivity toward ingroup members, it should be noted that greater mentalizing, as indexed by enhanced TPJ activity, may not be solely reserved for ingroup members. For example, recent neuroimaging work found that White perceivers who were highly internally motivated to respond without prejudice (IMS; Plant & Devine, 1998) showed greater TPJ activity when watching videos of perceived Black people being arrested as the arrests became more aggressive relative to when they viewed similarly violent depictions of White people being arrested (Dang et al., 2022). This suggests that while the TPJ tends to show preferential activity toward ingroup members, at times, it may activate preferentially toward outgroup members. Additionally, mentalizing indexed by TPJ activity may be sensitive to perceived unfairness of treatment people receive, regardless of whether the person receiving unfair treatment is a member of their ingroup.

STS

In addition to the DMPFC and TPJ, the STS has also been implicated in mentalizing. The STS generally serves as a hub that communicates with brain regions associated with social perception and social cognition (Deen et al., 2015). The STS is implicated in gleaning inferences from others based on perceptual information (Cloutier, Turk, et al., 2008; Hoffman & Haxby, 2000; Schurz et al., 2014). To this end, the STS may not be as central to intergroup differences in mental state inferences as the TPJ and DMPFC. Nonetheless, recent work has found that increased bilateral STS activity was observed when White perceivers view aggressive arrests of perceived Black civilians (compared with perceived White civilians) by perceived White officers (Dang et al., 2022). Almost the entire sample of this study showed high levels of internal motivation to avoid prejudice, suggesting that participants' egalitarian worldviews may be associated with greater mentalizing toward perceived Black people who may have been unjustly arrested.

Summary: Mentalizing and Intergroup Bias

Taken as a whole, the TPJ, DMPFC, and STS comprise a network of brain regions that underlie various aspects of mentalizing. Reduced activity in all three regions, but particularly the DMPFC and TPJ, may be associated with greater degrees of ingroup preference, such that people who are members of outgroups may be afforded less mentalizing. Although greater activity in mentalizing regions has been observed in response to outgroup members receiving treatment that is believed to be

unfair (e.g., Dang et al., 2022), greater mentalizing generally occurs for ingroup members.

Regulation of Intergroup Bias

In most contexts, few are willing to openly express biased attitudes, even among those who explicitly hold them. Existing work has found that those motivated to regulate their prejudice will attempt to do so, even when it is inefficient (Ito et al., 2015; Mattan, Kubota, Dang, et al., 2018; Richeson et al., 2003). However, this may not be true for those with greater degrees of prejudice or who lack the motivation to regulate racial bias (e.g., Forscher et al., 2015; Schmader et al., 2022). A better understanding of the differences between people who are more externally motivated to regulate their biases and those who lack the motivation to regulate their prejudice could provide promising insights into the regulatory and control mechanisms associated with intergroup biases. To that end, the following section offers some background information on an exploratory set of analyses on regions implicated in the regulation of cognitive control, such as the dorsal anterior cingulate cortex (DACC), the dorsolateral prefrontal cortex (DLPFC), and the ventrolateral prefrontal cortex (VLPFC), help map how people may regulate their biases.

DLPFC

The DLPFC has been implicated in executive function and response inhibition (e.g., the conscious monitoring of responses and selection of appropriate responses

when biases and responses may be in conflict; Badre & Wagner, 2004; Cassidy et al., 2016; Hartley & Phelps, 2010; Richeson et al., 2003; Sharp et al., 2004; Sreenivasan et al., 2014). It is, therefore, not surprising that the DLPFC may be involved when intergroup bias needs to be regulated. For example, in an investigation by Cassidy et al., (2016), younger and older White adults viewed images of perceived White and perceived Black faces. The authors argued that because older age is associated with diminished cognitive control (Kwee & Nakada, 2003), the authors suggested that comparisons of brain activity in the DLPFC between the older and younger participants would reflect differences in cognitive control while viewing images of faces of Black and White people. Their results first found that older adults showed generally diminished activity of the DLPFC toward perceived Black faces. Additionally, they found that older adults who had diminished DLPFC activity toward perceived Black faces showed diminished functional connectivity between the DLPFC and amygdala. Therefore, the authors suggested that the DLPFC may be implicated in the application of cognitive control of negative evaluations of outgroup members.

DACC

The DACC is another region that has been implicated in behavioral regulation and conflict monitoring (Botvinick et al., 2001; Kolling et al., 2016). Work concerning intergroup biases has posited that the DACC is similarly involved in the regulation of intergroup bias in that the DACC is preferentially recruited during the detection of conflicts between one's own biases and motivations not to let bias influence behavior

(Amodio, 2014; Kubota et al., 2012; Mattan, Wei, et al., 2018) and has been suggested to hold similar functions as the DLPFC when regulating intergroup biases (Cassidy et al., 2016; Richeson et al., 2003). Because of the converging patterns of activity, it has been posited that the DLPFC may work in concert with the DACC during the regulation of racial bias (Amodio, 2014; Kubota et al., 2012; Mattan, Wei, et al., 2018) such that the DACC is involved in reconciling stereotypical knowledge accessible to a person with their motivations and goals. As such, the DACC plays a role in cognitive control related to the regulation of intergroup biases and can be preferentially recruited depending on one's motivation to regulate their responses.

VLPFC

The VLPFC is another region that is commonly associated with regulating negative affect (Ochsner et al., 2012; Wager et al., 2008) that has been shown to preferentially respond during the regulation of negative affect from threatening stimuli (Hartley & Phelps, 2010; Tupak et al., 2014). Relatedly, the VLPFC has been associated with impulse control (Aron et al., 2004, 2014). Diminished activity or damage to the VLPFC has been associated with changes in how rewarding stimuli are evaluated. Specifically, reduced VLPFC involvement appears to lead to smaller immediate rewards being preferred, even at the cost of larger rewards in the long term (Kim & Lee, 2011; Wagner et al., 2013). Beyond its role in affect regulation and appraisals of rewarding stimuli, there is limited evidence that implicates the VLPFC in

intergroup processing. More specifically, this limited body of work suggests the VLPFC and, more specifically, the IFG (Levy & Wagner, 2011) is involved in the regulation of intergroup biases (Amodio & Cikara, 2021; Mattan, Wei, et al., 2018; Richeson et al., 2003; Telzer et al., 2015). In sum, along with the DACC and DLPFC, the VLPFC is thought to be involved in the regulation of intergroup biases.

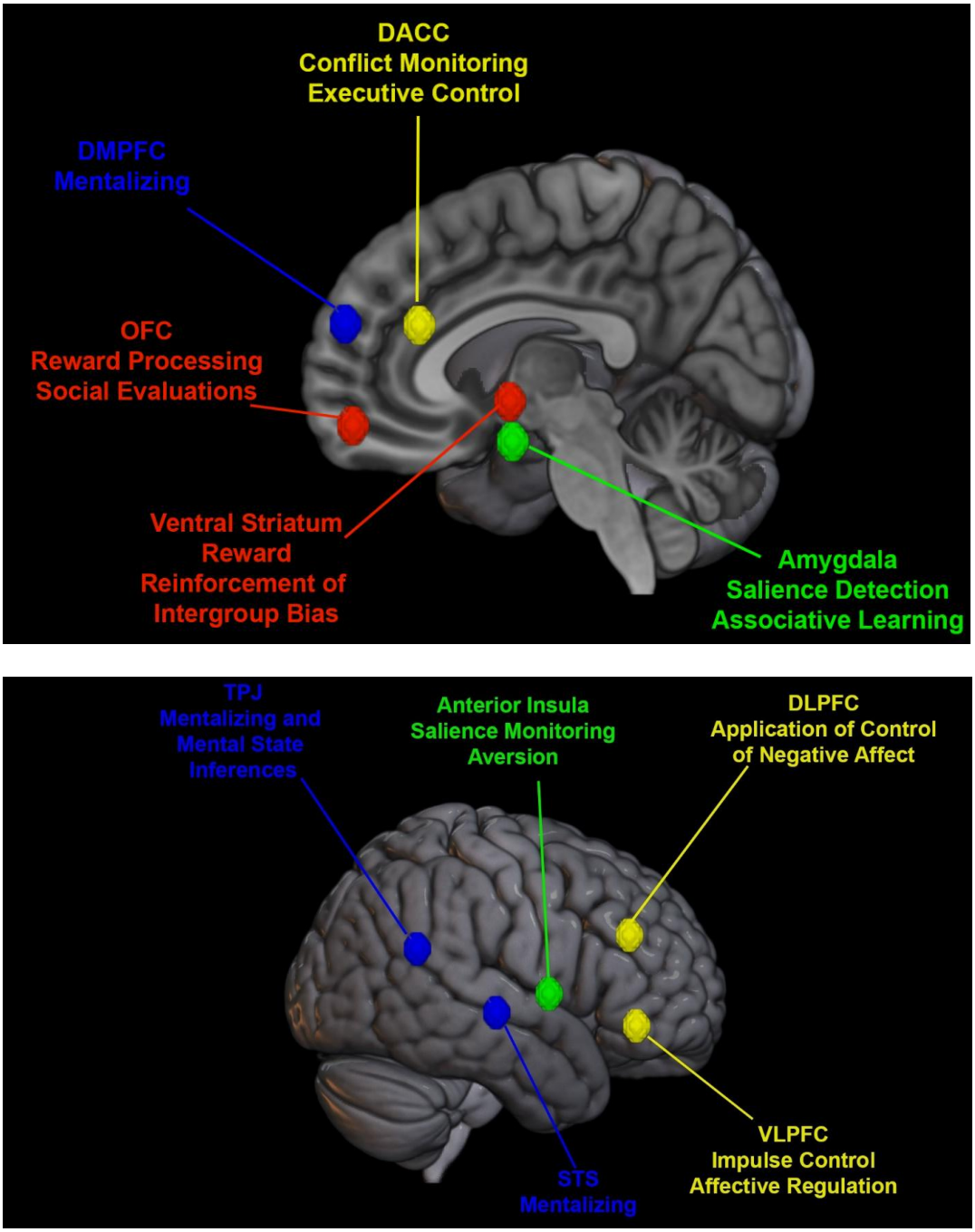

Figure 1. Brain regions implicated from neuroscientific investigations relevant to intergroup bias. The NACC/VS and OFC are implicated in reward and social learning, that can further facilitate intergroup biases. The amygdala and AI have been implicated in salience monitoring and detection. Moreover, the amygdala has been implicated in the learning of threat associations with perceived outgroup

members. The DMPFC, STS, and TPJ have been implicated in mentalizing, with decreasing activity being observed toward outgroup members or recipients of unfair treatment. The DLPFC, DACC, and VLPFC have been implicated in regulation, with the DACC being implicated in conflict monitoring and the DLPFC and VLPFC being implicated in the application of control with respect to intergroup prejudice and stereotyping. Further, the VLPFC has been associated with impulse and affect control. Regions supporting reward are colored in red, and regions that support salience are colored in green. Blue regions have been associated with mentalizing, and regions colored yellow have been associated with regulation.

Operationalizing Drivers of Intergroup Bias

Existing social neuroscientific work on intergroup dynamics has generally found that brain regions tend to activate in support of greater positive biases toward ingroup members and more negative biases toward outgroup members. Recall a common thread of social identity theory, which is the categorization of similar others as ingroup members and dissimilar others as outgroup members. This theoretical framework suggests that, for example, among White Americans, there may be a favoritism toward other White people. People with such strong ingroup preferences may glean a strong sense of self-worth from their membership to their perceived ingroup (Tajfel & Turner, 1986) and, in turn, ascribe highly positive appraisals to their ingroup (Brewer, 1999). Within existing social psychological frameworks, this positive appraisal, or ingroup love, is generally thought to be a path to intergroup discrimination (Brewer, 1979; Tajfel & Turner, 1979) that can influence intergroup prejudices and discrimination without the presence of outgroup animus (Brewer, 2017). Most social cognitive work that has examined intergroup dynamics has used

face stimuli rather than more complex depictions of social interactions, potentially missing nuances of how people extract and process information about the behaviors of ingroup and outgroup members. Because Social Identity Theory (Tajfel & Turner, 1986) is rooted in ingroup affiliation on one end and outgroup aversion on the other, leveraging depictions of social interactions that are either affiliative or aversive may provide richer insights into the neural underpinnings of intergroup processing than using face stimuli. Additional context provided by depictions of social behavior among ingroup and outgroup members may provide more nuance as to how perceivers engage brain regions supporting intergroup processing. To that end, depictions of affiliative interactions (i.e., cross-race dating) and aversive interactions (i.e., cross-race violence) may be uniquely suited to provide a more nuanced understanding of how White perceivers differentially process intergroup social interactions.

In summary, this dissertation will consist of three studies that examine intergroup processing among a sample of White Americans using stimuli that depict affiliative or aversive intergroup social interactions. In Study 1A, we discuss the development of an equated stimuli set depicting smiling couples who differ in racial composition. Study 1B similarly discusses the development and piloting of an additional equated set of images, depicting scenes of either brutal violence by a white aggressor toward perceived Black or perceived White victims or scenes depicting mundane social interactions.

Study 2 examined brain regions that we believed would be implicated in the processing of affiliative social interactions. We centered our main analyses on brain

regions that support positive social evaluations and those that support the processing of salience of outgroup members. Specifically, in Study 2, our confirmatory analyses examined differences in activity in the OFC, NACC, amygdala, and AI when participants viewed the images of couples comprised of different race partners. Additionally, we conducted exploratory analyses on regions that support the regulation of intergroup bias (i.e., the DACC, DLPFC, and VLPFC).

Lastly, Study 3 similarly examined brain regions we believed would be implicated in the processing of the aversive social interactions. Our analyses centered on brain regions supporting the processing of aversive stimuli and mentalizing. The confirmatory analyses of Study 3 focused on differences in the amygdala, AI, TPJ, and STS when participants processed violent images or images depicting mundane social interactions among perceived White men. Additionally, our exploratory analyses focused on regions that support the regulation of intergroup bias (i.e., the DACC, DLPFC, and VLPFC) as well as the DMPFC. We then conclude the dissertation with a general discussion of the three studies' findings within the context of existing theory in social psychology and social neuroscience, as well as future directions from the studies presented in this thesis.

Chapter 2

STUDY 1A: CURATION OF AFFILIATIVE INTERGROUP STIMULI

Within Social Identity Theory, ingroup affiliation is generally believed to be viewed as more positive than outgroup affiliation. However, it is unclear how depictions of intergroup affiliative behavior may be processed by White perceivers when individuals being depicted are of both an ingroup and an outgroup. As such, we opted to create a first set of stimuli to more accurately examine how perceivers process depictions of happy smiling couples who vary in racial composition. To that end, in Study 1A, we curated a set of normed stimuli depicting positive, affiliative intergroup social interactions. In the following section, we detail the steps taken to curate this first set of stimuli.

Methods

Image Curation

Potential photo stimuli were collected from across the internet. To assess the extent to which individuals evaluate different racial compositions of couples, photos depicting two perceived White partners, two perceived Black partners, or where one

partner was perceived to be Black, and one was perceived to be White. The stimuli curated in Study 1A were originally intended to be used in a series of studies that examined the psychological and neural underpinnings that may account for the escalation of explicit racial biases into the explicit endorsement of white supremacist worldviews. White supremacist worldviews as a construct encompass a variety of explicit biases based on different social identities (e.g., K. Harris et al., 2020; Reyna, Bellovary, et al., 2022; Reyna, Harris, et al., 2022). In addition to race-based biases, white supremacist beliefs tend to encompass patriarchal, misogynistic worldviews, where women are viewed as subservient to men (Blee, 2002; Latif et al., 2020; McMahon & Kahn, 2018; Pruden et al., 2022; Stillwell & Lowery, 2021). Moreover, within white supremacist extremist groups, White women are viewed as a means to preserve the "White race" through childrearing (Latif et al., 2020; McMahon & Kahn, 2018) and depictions of White women being violated have been widely used in white supremacist propaganda (McMahon & Kahn, 2018; Pruden et al., 2022). Therefore, because of the visceral nature of depictions of White women and Black men being used as a rallying cry by white supremacists, we only opted to collect images of Interracial couples where the perceived man appeared to be Black, and the perceived woman appeared to be White. In total, 299 photos of couples were collected from the internet for piloting ($n_{Blackcouples}$ = 103, $n_{Interracialcouples}$ = 89, $n_{Whitecouples}$ = 107).

Because the images depicted complex social interactions between happy smiling couples, we wanted to ensure that perceivers were processing social information about the couples that reflect ingroup or outgroup identities rather than

potentially confounding information (e.g., if one couple race composition appeared more or less authentic than the other two). As such, we tasked independent samples of participants to rate the images on a number of dimensions that we believed would help avoid confounds in the images across couple race conditions. Prior to collecting image ratings, each image was cropped into a square image, greyscaled, and resized to a 600 x 600-pixel size.

All images were rated across three round of ratings. The decision to split ratings across three rounds was first made to help reduce participant dropout, as without splitting up rounds of ratings, participants may have needed to potentially complete over 1,000 ratings. Additionally, in addition to collecting ratings about the entire images, we also collected ratings about either individual in the image. To allow participants to assess each individual depicted in the scenes, red arrows were superimposed in each image using Adobe Photoshop (https://www.adobe.com/) that pointed to every individual within the photo. During the rounds of rating that involved assessing the individual within each scene, participants viewed the images with the superimposed arrows pointing to each person in the scene rather than viewing and rating the whole scene itself (see Figure 2). As such, it was necessary to collect ratings across three rounds and, in the second and third rounds, split up ratings such that one sample of raters rated the entire image, and another sample rated the individuals within each image.

The image ratings were collected using Inquisit 5.0 (Millisecond.com). During the image rating tasks, participants were shown the images of couples in random

presentation order. Participants were instructed to respond to each rating dimension using their keyboard to indicate their responses. Two steps were taken to ensure attentiveness to the rating task and deliberate responses from participants. First, participants were instructed to make deliberate ratings for each image. However, participants were also told to slow down their ratings if they responded too quickly (i.e., < 300ms post-image presentation) or to make their decisions more quickly if their responses were too slow (i.e., < 10,000ms). These prompts remained on screen for 2000ms to dissuade participants from repeatedly rushing through trials. In addition to the prompts for responses that were too fast or too slow, participants were tasked with a simple attention check that appeared approximately after every 36 images. Participants were initially instructed to actively monitor for the attention checks during the initial task instructions. The attention checks consisted of an image of a couple that was not being rated with the other stimuli. Superimposed on each of the attention check images were numbers one through eight (see Figure 3). When the attention check prompt appeared rather than a rating prompt, participants were told to press the corresponding number that appeared in the image on their keyboard to show they were paying attention to the ratings. Participants who did not correctly pass enough of the attention checks had their ratings excluded from the final equating.

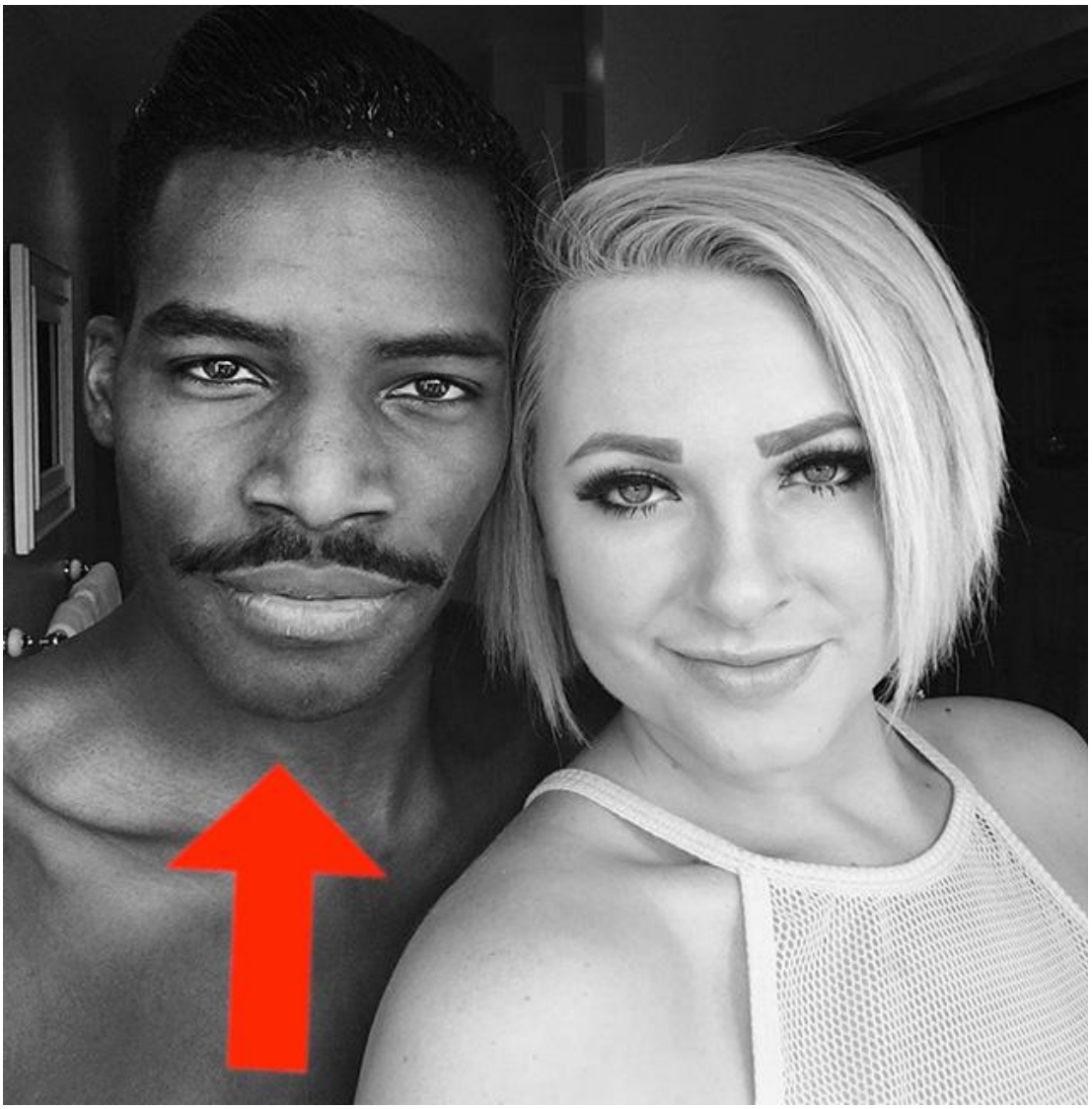

Figure 2. Example images of the superimposed red arrow pointing to either partner to aid with the ratings for either partner in the scene.

Across all rounds of Study 1A, all participants provided informed consent in accordance with the guidelines set by the Declaration of Helsinki, and the study protocol was approved by the University of Delaware's Institutional Review Board.

Figure 3. Example of image used for the attention checks for the couples rating task. Each attention check image had a number between one and eight superimposed in the scene.

Rating Round 1

An independent sample of 59 participants was recruited from Amazon Mechanical Turk (MTurk) to rate each photo. Three participants were excluded for not completing the ratings entirely, and another six participants were excluded for failing to successfully pass at least seven out of nine attention checks. As such, the final sample of raters for the first round consisted of 50 participants from MTurk (24 women, 24 men, 1 preferred to self-describe, 1 did not report gender; M_{age} = 37.48 SD_{age} = 10.585; 39 White/ European American, 8 Black/African American, 1 Multiracial, 1 Latinx, 1 Native American). These individuals rated all 299 photos on how high quality they perceived the images to be, as well as the time of day that they believed the photo to have taken place. Quality ratings were assessed using a 1-7 Likert scale ranging from "*1-Very low quality*" to "*7-Very high quality*". Participant's quality ratings were averaged per photo and were used as an initial exclusion criterion. Images that had an average quality rating of less than 3/7 were excluded ($n_{Blackcouples}$ = 8, $n_{Interracialcouples}$ = 11, $n_{Whitecouples}$ = 9), leaving a total of 271 photos.

Rating Round 2

A sample of 60 participants recruited from MTurk rated each individual within the photos of couples on whether they perceived either partner as being a man or woman with (i.e., *1 = "yes", 7 = "no"*) as well as what the orientation of each person's gaze was (i.e., *1 = "direct", 4 = "indirect", or 7 = "unclear"*). Eye gaze

was not an exclusion criterion for the second round of ratings. However, we wanted to control for it in light of some cognitive neuroscientific findings that suggest some regions supporting social cognition may activate differently depending on gaze orientation (e.g., Engell & Haxby, 2007). Each participant viewed the images of couples where a red arrow was superimposed in the image that pointed to either person in the scene to indicate who they were rating. Two participants were excluded from the second round of ratings due to attention check failures (i.e., failing more than 3/15 attention checks). Additionally, another six participants were excluded for having incomplete datasets, leaving a sample of 52 raters (20 women, 31 men, 1 did not report gender; M_{age} = 36.48 SD_{age} = 9.11; 36 White/ European American, 5 Black/African American, 5 Asian American/ Pacific Islander, 4 Latinx, 2 Native American) for the second round of stimuli ratings. Gender consensus ratings were calculated for either partner, and images containing at least one partner who had less than 80% consensus were excluded. Only one image from the perceived White couples' condition was excluded for low consensus on perceived gender in round 2, leaving a total of 270 potential images.

Rating Round 3

An independent sample of 66 participants recruited from MTurk rated the image on several dimensions. Three participants did not successfully pass six out of eight attention checks in the task, and an additional 13 participants did not complete

the ratings in their entirety. Following exclusions, 50 participants (33 women, 17 men; M_{age} = 38.92 SD_{age} = 11.025; 38 White/ European American, 4 Black/African American, 1 Asian American/ Pacific Islander, 2 Multiracial, 4 Latinx, 1 preferred to self-describe) rated the images on whether they were familiar to participants (i.e.,"*1-Yes*", *"2-No"*), how authentic they believed the couple to be (i.e.,"*1-Not at all authentic*" to "*7-Completely authentic*"), and how romantically engaged they couple appeared to be (i.e.,"*1-Not at all romantically engaged*" to "*7-Completely romantically engaged*"). Additionally, these participants were asked what decade they believed the photo to have been taken in (i.e., *1 = 2020s (2020-2021), 2 = 2010s (2010-2019), 3 = 2000s (2000-2009), 4 = 1990s (1990-1999), 5 = 1980s (1980-1989), 6 = 1970s (1970-1979), 7 =1960s (1960-1969) 8 = 1959 or before*).

An additional sample of 69 participants rated the individuals within the image on their perceived age (i.e., *1 = "18-25 years old"; 2 = "26-35"; 3 = "36-45"; 4 = "46-55"; 5 = "56-65"; 6 = "66-75";7 = "76+ years old"*), perceived race (i.e., *1= Black/African American, 4 = White/Euro American, 7 = Other/Not Clear*), and the valence of perceived emotional expressions (i.e.,"*1-Very Negative* " to "*7-Very Positive*"). Seventeen participants were excluded for failing to complete the ratings in their entirety. An additional two participants were excluded for failing to pass attention checks with at least 80% accuracy. As such, the final sample of raters consisted of 50 participants (28 women, 22 men; M_{age} = 39.9 SD_{age} = 9.59; 39 White/ European American, 2 Black/African American, 1 Asian American/ Pacific Islander, 5 Multiracial, 3 Latinx) who rated either individual within the scenes. Race consensus

scores were calculated for either individual in the photo and images that contained at least one partner with low race consensus (i.e., consensus less than 75%) were excluded. A total of 38 images were excluded in round 3 for low consensus on the perceived race of each partner ($n_{Blackcouples}$ = 5, $n_{Interracialcouples}$ = 9, $n_{Whitecouples}$ = 24), leaving a total of 232 images to be equated.

Results

Image Equating

Images ratings were analyzed using ANOVAs to test for differences in ratings across race conditions using the EZ package (Lawrence & Lawrence, 2016) in the R programming language (R Core Team, 2017; see Table 1 for all dimensions equated and their inferential statistics; see Table 2 for the descriptive statistics for the couples stimuli). Analyses resulted in three sets of 20 photos that were equated across nine dimensions.

Couples were first equated as being similarly perceived as highly authentic across couple race. The photos were on average, believed to have taken place in the 2010s, were rated as being equally unfamiliar, and the couples were rated as being equally romantically engaged across all three couple race conditions. The women in all photos were equated such that they were all rated as being similar in perceived age, and the valence of their facial expressions was rated as being similarly positive. Similarly, the men in all photos were equated such that they were also rated as being

similar in perceived age, and their emotional expressions were also similarly rated as equally positive. Lastly, the equated images were equated on luminance and contrast using the SHINEToolbox in MATLAB (Willenbockel et al., 2010).

Table 1 Stimuli Equating Results Across Perceived Black, Interracial, and Perceived White Couples

Rating Dimension	***F***	***df*$_1$**	***df*$_2$**	***p***
Perceived authenticity	0.190	2	57	0.827
Perceived decade	0.497	2	57	0.611
Familiarity	0.557	2	57	0.576
Perceived romantic engagement	0.508	2	57	0.604
Women's perceived valence of emotional expression	0.529	2	57	0.592
Women's perceived age	1.508	2	57	0.230

Men's perceived valence of emotional expression	0.610	2	57	0.547
Men's perceived age	1.572	2	57	0.217
Proportion Black Women	15065.434	2	57	< .001
Proportion White Women	4816.170	2	57	< .001
Proportion Black Men	8036.338	2	57	< .001
Proportion White Men	11833.688	2	57	< .001

Results from the final equated set of images of couples across all three racial compositions. The final set of 60 images were equated on how authentic the couples appeared to be, the decade the images were believed to have taken place in, how familiar the image was to the sample of raters, how romantically engaged the couples appeared to be, perceived age and intensity of emotional expression for both men and women for perceived Black, Interracial and perceived White couples. Moreover, the proportion of raters reporting that they were not familiar with the images did not significantly differ as a function of couple racial composition.

Table 2 Descriptive Statistics for Each Condition of Couple Race

Rating Dimension	M_{Black}	SD_{Black}	$M_{Interracial}$	$SD_{Interracial}$	M_{White}	SD_{White}
Perceived authenticity	5.154	0.248	5.124	0.236	5.103	0.300
Perceived decade	2.347	0.185	2.269	0.286	2.364	0.440
Familiarity	6.946	0.061	6.964	0.069	6.964	0.056
Perceived romantic engagement	5.138	0.255	5.175	0.468	5.068	0.253
Proportion of unfamiliarity with image	0.991	0.010	0.994	0.011	0.994	0.009
Women's perceived valence of emotional expression	5.972	0.229	5.895	0.387	6.011	0.439
Women's perceived age	2.401	0.557	2.343	0.367	2.674	0.894
Men's perceived valence of emotional expression	5.918	0.516	5.715	0.637	5.792	0.601

Men's perceived age	2.528	0.590	2.409	0.392	2.788	0.965
Proportion Black Women	0.981	0.023	0.021	0.024	0.009	0.012
Proportion White Women	0.009	0.012	0.947	0.049	0.972	0.035
Proportion Black Men	0.994	0.011	0.965	0.045	0.013	0.013
Proportion White Men	0.005	0.009	0.015	0.019	0.972	0.033

Descriptive statistics from the final equated set of images of couples across all three racial compositions.

Conclusion

In Study 1A, we compiled, piloted, and equated a set of images that depict happy, romantically engaged individuals that were comprised of a perceived Black man and a perceived Black woman, a perceived Black Man and a perceived White Woman, or a perceived White man and a perceived White woman. Because of the affiliative nature of the social interactions depicted in the scenes, these images may be used to examine how differences in the racial composition of those in affiliative social

interactions may impact neural activity supporting social evaluation, salience, and aversion.

Chapter 3

STUDY 1B: CURATION OF AVERSIVE INTERGROUP STIMULI

In addition to being related to greater positive appraisals of ingroup members engaging in positive social interactions, Social Identity Theory suggests that depictions of aversive social interactions toward ingroup members may be viewed as aversive relative to the recipient of the aversive behavior was an outgroup member. However, it is not immediately clear how perceivers may process depictions of aversive social interactions involving harm to ingroup members relative to depictions where an outgroup member is being harmed or when there are no such perceptions of harm. To that end, we opted to create a second set of stimuli to more accurately examine how perceivers process depictions of violence when the victim is either in their ingroup or an outgroup. To that end, in Study 1B, we curated a set of normed stimuli depicting negative, aversive intergroup social interactions, namely, depictions of intergroup violence. In the following section, we detail the steps taken to curate this second set of stimuli.

Methods

Violent Image Curation

Potential photo stimuli were collected from across the internet. To assess the extent to which individuals differentially evaluate depictions of violence depending on the perceived race of the victim, photos depicting a White aggressor or aggressors and either perceived Black or perceived White victim(s). In total, 174 photos of violent scenes were collected for piloting ($n_{Blackvictim}$ = 85, $n_{Whitevictim}$ = 89). Similar to the image ratings in Study 1A, we wanted to control for potential confounds in the violent stimuli to help ensure that participants were attending to the perceived race of the victims rather than other potential confounding variables cross stimuli condition (e.g., if the facial expression of one violent condition was more or less negative than the other). Prior to collecting image ratings, every image was cropped into a square image and resized to a 600 x 600-pixel size.

Mirroring Study 1A, all images were rated across three rounds of ratings. The decision to split ratings across three rounds was again made to help reduce participant dropout, as without splitting up rounds of ratings, participants may have needed to potentially complete over 1,000 ratings. Additionally, in addition to collecting ratings about the entire image, we also collected ratings about all individuals in the image. To allow participants to assess each individual depicted in the scenes, red arrows were superimposed within the photo and pointed to each individual. During the rounds of

rating that involved assessing the individual within each scene, participants viewed images with arrows superimposed in the image using Adobe Photoshop (https://www.adobe.com/). The arrows pointed to each person in the scene. Participants were either tasked with rating the individuals in the scene or rated the whole scene itself (see Figure 4) across several dimensions. As such, it was necessary to collect ratings across three rounds, and in the second and third rounds, split up ratings such that one sample of raters rated the entire image, and another sample rated the individuals within each image.

Figure 4. Example images of the superimposed red arrow pointing to each individual in the scene to aid with the ratings for everyone in the foreground of the scene.

The image ratings were collected using Inquisit 5.0 (Millisecond.com). During the image rating tasks, participants were shown the images of violence in random presentation order. Participants were instructed to respond to each rating dimension using their keyboard to indicate their responses. Two steps were taken to ensure attentiveness to the rating task and deliberate responses from participants. First,

participants were instructed to make deliberate ratings for each image. However, participants were also told to slow down their ratings if they responded too quickly (i.e., < 300ms post-image presentation) or to make their decisions more quickly if their responses were too slow (i.e., < 10,000ms). These prompts remained on screen for 2000ms to dissuade participants from repeatedly rushing through trials. In addition to the prompts for responses that were too fast or too slow, participants were tasked with a simple attention check that appeared approximately after every 36 images. Just like in Study 1A, participants were initially instructed to actively monitor for the attention checks during the initial task instructions. The attention checks consisted of an image that was not being rated with the other stimuli. Superimposed on each of the attention check images were numbers one through eight (see Figure 5). When the attention check prompt appeared rather than a rating prompt, participants were told to press the corresponding number that appeared in the image on their keyboard to show they were paying attention to the ratings. Participants who did not correctly pass enough of the attention checks had their ratings excluded from the final equating.

Figure 5. Example of an image used for the attention checks for the violent image rating task. Each attention check image had a number between one and eight superimposed in the scene.

Across all rounds of Study 1A, all participants provided informed consent in accordance with the guidelines set by the Declaration of Helsinki, and the study protocol was approved by the University of Delaware's Institutional Review Board.

Rating Round 1

A sample of 55 participants was recruited from MTurk using TurkPrime. Four participants were excluded for failure to complete the full rating task. All remaining participants completed the rating task while also successfully completing five out of seven attention checks. The resulting sample consisted of 51 independent raters (25 men, 26 women; $M_{age} = 45.5$ $SD_{age} = 13.4$; *35* White/European American, 2 Asian American/ Pacific Islander, 9 African American/Black, 5 Latinx). The sample of raters evaluated each image on perceived image quality on a 1-7 Likert scale (i.e.,"*1-Very low quality*" to "*7-Very high quality*"), time of day (i.e., *1 = Day, 4 = Night, 7 = Unsure*), and whether there was a clear attacker and a clear victim (i.e., *1 = Yes, 7 = No)*. Consensus ratings of at least 80% of participants agreeing that each image depicted both a clear attacker(s) or victim(s) was required for images to move to subsequent rounds of ratings. A total of 69 images were excluded due to being rated as low quality ($n_{Blackvictim} = 38$, $n_{Whitevictim} = 31$). Leaving a total of 105 images to be rated in round 2.

Rating Round 2

Two rounds of raters were recruited during the second round of ratings for the violent images. The decision to split the second round of rating into two waves was first made to similarly help reduce participant attrition from too many trials in the rating tasks. Second, because there were potentially more than three individuals depicted in each scene, we wanted to ensure that participants were only focusing on the victims or aggressors to avoid any interference from having to switch between rating the victims and aggressors. In the first wave of ratings, participants rated only the victims present in the scene, while in the second wave, they only rated the aggressors in the scene. Participants were tasked with responding to whether they thought each victim or aggressor was a man or woman using categorical responses (i.e., *1 = "Yes", 4 = "No", 7 = "Unsure"*). Additionally, they rated each victim or aggressor based on their gaze orientation (i.e., *1 = "Direct", 4 = "Indirect", 7 = "Unsure"*). Either victim(s) or aggressor(s) were highlighted by a large red arrow superimposed over the person in the scene. Similar to Study 1A, eye gaze was not an exclusion criteria for the second round of ratings. However, we wanted to control for it in light of some cognitive neuroscientific findings that suggest some regions supporting social cognition may activate differently depending on gaze orientation (e.g., Engell & Haxby, 2007).

A sample of 64 participants was recruited from MTurk using Turk Prime for the victim ratings. Ten were excluded for failing to respond to accurately to at least six

out of eight attention checks. An additional four participants were excluded for stopping the task part way through. Leaving a final sample of 49 participants (26 men, 22 women; $M_{age} = 38.98$ $SD_{age} = 11.15$; 38 White/European American, 4 Native American, 5 African American/Black, 1 Latinx, 1 individual did not report demographics).

An additional 63 participants were recruited from MTurk using TurkPrime for the aggressor ratings. Seven participants were excluded for failing to respond to accurately to at least six out of eight attention checks. An additional two participants were excluded for failure to complete the task in its entirety. Leaving a final sample of 54 participants (21 men, 30 women; $M_{age} = 36.98$ $SD_{age} = 10.94$; 34 White/European American, 4 Asian American/ Pacific Islander, 6 African American/Black, 3 Multiracial, 4 Latinx, 3 individuals did not report demographics).

Consensus ratings of at least 80% of participants about the perceived gender of each person in the scene was a man was required for images to move to the third round of ratings. Nine images were excluded in the second round of ratings due to low consensus about the gender of people depicted in the scene ($n_{Blackvictim} = 6$, $n_{Whitevictim} = 3$), leaving a total of 96 images to be rated in the third and final round of rating.

Rating Round 3

Similar to round two, the third round of ratings consisted of two groups of independent raters. The first sample of raters evaluated the individuals within the

scene on several dimensions. The person in the scene that they were rating was indicated by a large red arrow superimposed in the photo that pointed to the person they were rating. A second sample of participants evaluated each scene on several dimensions.

Fifty participants were recruited from MTurk using TurkPrime for the person-level ratings. Zero participants had less than 75% accuracy on the 11 attention checks, and no participants quit the rating task partway through, leaving a final sample of 50 participants (23 men, 27 women; $M_{age} = 38.46$, $SD_{age} = 9.09$; 37 White/European American, 3 African American/Black, 3 Asian American/ Pacific Islander, 2 Multiracial, 2 Latinx, 1 Middle Eastern/ Arab American). Raters evaluated each aggressor and victim in the scene on their perceived age (i.e., *1 = "18-25 years old"; 2 = "26-35"; 3 = "36-45"; 4 = "46-55"; 5 = "56-65"; 6 = "66-75";7 = "76+ years old*) and the valence of emotional expression on a 1-7 Likert scale (i.e., *"1 = very negative" to "7 = very positive"*). Additionally, consensus ratings were taken for the perceived race of each person in the scene (i.e., *1 = "Black/ African American", 4 = "White/Euro American", 7 = "Other/Unsure").* Lastly, participants rated each person in the scene on whether they believed them to be police or law enforcement (i.e., *1 = "Yes", 4 = "No", 7 = "Unsure").*

An additional 50 participants were recruited from MTurk using TurkPrime for the scene-level ratings. Similarly to the person-level ratings, zero participants had less than 75% accuracy on the seven attention checks, and no participants quit the rating task partway through, leaving a final sample of 50 participants (17 men, 31 women, 1

Non-Binary, 1 Transgender; M_{age} = *39.86,* SD_{age} = *12*.80; 38 White/European American, 6 African American/Black, 5 Latinx, 1 Asian American/ Pacific Islander). Participants rated each scene on how violent they believed the scene was on a 1-7 Likert scale (i.e., *1 = "Not at all", 7 = "Extremely"*), the decade they believed it to have taken place in (i.e., *1 = 2020s (2020-2021), 2 = 2010s (2010-2019), 3 = 2000s (2000-2009), 4 = 1990s (1990-1999), 5 = 1980s (1980-1989), 6 = 1970s (1970-1979), 7 =1960s (1960-1969) 8 = 1959 or before*), and the likelihood that the scene occurred within the United States on a 1-7 Likert scale *(i.e., "1 = Not at all likely" 7 = "Very likely")*. Participants also provided consensus ratings on whether they were familiar with the photo (i.e., *1 = "yes", 7 = "no"*), and whether they believed that anyone was actively being harmed in the photo (i.e., *1 = "yes", 7 = "no"*).

Consensus ratings on the perceived race of each person in the scene and familiarity ratings were used as cutoff criteria for the third round of ratings. Images with less than 80% consensus on there being someone actively being harmed in each photo were excluded. Additionally, images where less than 80% of raters reported having never seen the image prior to rating were excluded. As a result, 30 images were excluded in the second round of ratings ($n_{Blackvictim}$ = 9, $n_{Whitevictim}$ = 21), leaving a total of 66 images to be equated across either condition.

Rating Round 1

A sample of 52 participants was recruited from MTurk using TurkPrime. One participant was excluded for not completing the eight attention checks during the rating task with at least 75% accuracy, and one participant did not complete the rating task entirely. As such, the final sample of raters for the first round of neutral image ratings consisted of 50 raters (27 men, 23 women; $M_{age} = 38.12$, $SD_{age} = 9.24$; 33 White/European American, 6 African American/Black, 5 Asian American/ Pacific Islander, 4 Latinx, 2 Multiracial). Participants rated each image on the perceived image quality using a 1-7 Likert scale (i.e., "*1-Very low quality*" to "*7-Very high quality*") and whether they believed there was either an aggressor or victim present in the scene (i.e., *1 = "Yes", 7 = "No"*). Participants also rated the images on the perceived time of day (i.e., *1 = "Day", 4 = "Night", 7 = "Unsure"*). Images were excluded if their average image quality scores were rated as a three out of seven or lower or if there was less than 80% consensus on whether there was a clear attacker or victim present in the scene. Three images were excluded for low consensus ratings on both the victim and attacker being present in the image. Additionally, a duplicate of one photo was included in the first round of rating by mistake and was removed from subsequent rounds, leaving a total of 131 images to be rated in round two.

Rating Round 2

A sample of 54 participants was recruited to rate each person in the neutral images on whether they believed the person to be a man or a woman (i.e., *1 = "Yes", 4 = "No", 7 = "Unsure"*). and perceived gaze orientation (i.e., *1 = "Direct", 4 = "Indirect", 7 = "Unsure"*). Three participants were excluded for failing to pass more than 75% of eight attention checks within the rating task, leaving a final sample of 51 participants (26 men, 25 women; M_{age} = 40.88, SD_{age} = 12.83; 39 White/European American, 4 Asian American/ Pacific Islander, 3 African American/Black, 2 Latinx, 2 Multiracial, 1 Middle Eastern/Arab American). Images with lower than 80% gender consensus ratings were excluded from the final round of image ratings. Only one image failed to meet the 80% perceived gender consensus cutoff and was excluded, leaving a total of 130 images to be rated.

Rating Round 3

Similar to the violent images, two independent samples of participants rated either the people in the scenes or the scenes on several dimensions.

A sample of 53 participants was recruited from MTurk using TurkPrime to rate each neutral image on several dimensions. One participant did not complete the rating task in its entirety, and one participant did not complete the eight attention checks with at least 75% accuracy. This left a final sample of 51 participants (26 men, 25 women; M_{age} = 35.24, SD_{age} = 8.98; 38 White/European American, 5 African American/Black,

5 Latinx, 2 Multiracial, 1 Asian American/ Pacific Islander). Participants rated each scene on the perceived decade they believe the scene to have taken place in (i.e., *1 = 2020s (2020-2021), 2 = 2010s (2010-2019), 3 = 2000s (2000-2009), 4 = 1990s (1990-1999), 5 = 1980s (1980-1989), 6 = 1970s (1970-1979), 7 =1960s (1960-1969) 8 = 1959 or before*), if the scene was familiar to them (i.e., *1 = "yes", 7 = "no"*), whether they believed that someone was actively being harmed in the scene (i.e., *1 = "yes", 7 = "no"*), how violent they believed the scene was (i.e., *1 = "Not at all", 7 = "Extremely"*), and how likely they believed it was that the scene took place in the United States on a 1-7 Likert scale (i.e., *"1 = Not at all likely" 7 = "Very likely"*).

An additional sample of 50 participants were recruited from MTurk using TurkPrime (22 men, 27 women, 1 Nonbinary; M_{age} = 40.88, SD_{age} = 12.83; 32 White/European American, 9 African American/Black, 5 Latinx, 2 Asian American/ Pacific Islander, 2 Multiracial). No participants were excluded for failure to meet completion criteria or failure to pass the 11 attention checks on the rating task with 75% accuracy. A total of 50 participants rated each person within each image on perceived emotional expression using a 1-7 Likert scale (i.e., *"1 = very negative" to "7 = very positive"*), perceived age of each person in the image, assessed in bins of approximately 10 years (i.e., *1 = "18-25 years old"; 2 = "26-35"; 3 = "36-45"; 4 = "46-55"; 5 = "56-65"; 6 = "66-75";7 = "76+ years old*). Whether each person in the image was believed to be law enforcement (i.e., *1 = "yes", 7 = "no"*) and the perceived race of each person in the image (i.e., *1 = "Black/ African American", 4 = "White/Euro American", 7 = "Other/Unsure")*, a total of 22 images were excluded

for low consensus on whether the individuals in the scenes were perceived as White (i.e., less than 80% of raters did not believe that all people in the scene were perceived as White). To avoid potential confounds with very positive or negative emotions being displayed by the individuals in each scene, we also decided to exclude neutral scenes where the average emotional valence rating (i.e., *"1 = very negative" to "7 = very positive"*), was either rated as very negative (i.e., a two out of seven or lower) or very positive (i.e., a five out of seven on emotional expression or higher). As such, an additional 24 images were removed for being rated as having individuals in the scenes showing relatively more intense emotional expressions. Following exclusions, a total of 84 images were eligible for equating.

Results

Image Equating

Image ratings were analyzed using ANOVA to test for differences in ratings across victim race conditions using the EZ package (Lawrence & Lawrence, 2016) in the R programming language (R Core Team, 2017). Because the images were rated by two completely separate waves of independent raters, pairwise comparisons between all three image conditions were used to determine equating. First, the two violent conditions were contrasted (i.e., perceived Black victims vs. perceived White victims) to test for differences across stimuli conditions. Next, either violent condition was contrasted against the neutral condition. These analyses resulted in three equated sets

of 20 photos (see Table 3 for the equating between the images depicting White and Black victims; see Table 4 for the equating between images depicting Black victims and the neutral scenes; see Table 5 for the equating between images depicting White victims and the neutral scenes; see Table 6 for the descriptive statistics for the two violent photo conditions and the neutral images).

The equated set of images depicting perceived Black and White victims were rated as having equivalent proportions of people actively being harmed in each scene and were rated as being equivalently unfamiliar to participants. The degree of how violent each scene was rated as being equivalent, and the images were generally believed to have occurred between the 1980's and 1990's. Furthermore, the aggressors were rated as being equally perceived as military or law enforcement, their perceived age was not rated as different across conditions, and the valence of their perceived emotional expressions was rated as being equivalent across victim conditions. Finally, victims were also similarly rated as being equally unlikely to be members of law enforcement or military, their perceived age and valence of their emotional expressions were viewed equally as negative across victim conditions.

Table 3 Violence image equating for images depicting perceived Black Victim(s) vs perceived White Victim(s)

Rating Dimension	***F***	***df_1***	***df_2***	***p***
Proportion of someone in the scene actively being harmed	2.697	1	38	0.109

Proportion of raters who indicated scene was not familiar to them	0.147	1	38	0.703
Likelihood scene occurred in The United States	26.403	1	38	< 0.001
How violent the scene was	0.365	1	38	0.549
The decade scene occurred	1.324	1	38	0.257
Proportion of aggressors who were law enforcement or military	2.637	1	38	0.113
Proportion of victims who were law enforcement or military	1.727	1	38	0.197
Proportion of aggressors who were perceived to be neither Black nor White	2.848	1	38	0.100
Proportion of aggressors who were perceived to be White	1.483	1	38	0.231
Proportion of aggressors who were perceived to be Black	0.906	1	38	0.347
Proportion of victims who were perceived to be neither Black nor White	338.136	1	38	< 0.001
Proportion of victims who were perceived to be White	338.136	1	38	< 0.001
Proportion of victims who were perceived to be Black	1217.892	1	38	< 0.001
Aggressor's perceived valence of emotional expression	1.483	1	38	0.231
Aggressor's perceived age	0.635	1	38	0.431

Victim's perceived valence of emotional expression	0.523	1	38	0.474
Victim's perceived age	3.431	1	38	0.072

Results from the final equated set of images of violent images depicting perceived Black victim(s) versus perceived White victim(s). The final set of 40 images was equated on how the proportion of raters who indicated someone was actively being harmed in the photo, the proportion of raters reporting that they were not familiar with the images, the decade the images were believed to have taken place in, the proportion of images in either category that where the aggressor or aggressors were law enforcement or military, the proportion of images in either category that where the victim or victims were law enforcement or military, the proportion of raters who rated the aggressors as perceived White, as well as the intensity and emotional expression of both aggressor(s) and victim(s).

Neutral images were then equated with images that depicted perceived Black victims on several dimensions. First, both the neutral scenes and those depicting Black victims were rated as being equally unfamiliar and were believed to have occurred between the 1980's and 1990's. Additionally, the individuals in the scenes were rated as being equal in perceived age. Moreover, the images depicting Black victims were rated as depicting someone actively being harmed more than the neutral images, and the emotional expression of people in the scenes depicting Black victims was rated as being significantly more negative.

Table 4 Violence image equating for images depicting perceived Black Victim(s) vs Neutral Scenes

Rating Dimension	***F***	***df*$_1$**	***df*$_2$**	***p***

Proportion of someone in the scene actively being harmed	49.811	1	38	< 0.001
Proportion of raters who indicated scene was not familiar to them	2.593	1	38	0.116
Likelihood the scene occurred in The United States	4.938	1	38	0.032
How violent the scene was	323.785	1	38	< 0.001
The decade scene occurred	1.268	1	38	0.267
Proportion of people in the image who were law enforcement or military	97.035	1	38	< 0.001
Proportion of people in the image who were perceived to be neither Black nor White	2.513	1	38	0.121
Proportion of people in the image who were perceived to be White	1361.625	1	38	< 0.001
Proportion of people in the image who were perceived to be Black	4831.535	1	38	< 0.001
Perceived valence of emotional expression of people in the image	145.635	1	38	< 0.001
Perceived age of people in the image	0.188	1	38	0.667

Results found that the neutral images were rated as being equally familiar, occurred in the same decade, and individuals in the images were viewed as equal in perceived age compared to the images depicting perceived Black victim(s). Additionally, the neutral images were rated as being significantly less violent and less likely to take place in the United States compared to the images with perceived Black victim(s). Lastly, the individuals in the neutral images were rated as having more positive emotional expressions than those in the images with perceived Black victim(s).

Lastly, neutral images were tested against images that depicted perceived White victims. First, both the neutral scenes and those depicting White victims were rated as being equally unfamiliar and were also believed to have occurred between the 1980's and 1990's. Additionally, the individuals in the scenes were rated as being equal in perceived age. Moreover, the images depicting White victims were rated as depicting someone actively being harmed more than the neutral images, and the emotional expression of people in the scenes depicting White victims was rated as being significantly more negative. In summary, both the scenes depicting Black victims and those depicting White victims were rated as equally unfamiliar and were believed to have occurred in the same decade as the neutral images. At the person-level ratings, the individuals in both scenes depicting perceived Black and White victims also did not significantly differ in perceived age relative to the neutral images.

Following equating across victim race conditions, the equated images were further equated on luminance and contrast using the SHINEToolbox in MATLAB (Willenbockel et al., 2010).

Table 5 Violence image equating for images depicting perceived White Victim(s) vs Neutral Scenes

Rating Dimension	*F*	df_1	df_2	*p*
Proportion of someone in the scene actively being harmed	16.644	1	38	< 0.001
Proportion of raters who indicated the scene was not familiar to them	1.406	1	38	0.243
Likelihood the scene occurred in The United States	16.36	1	38	< 0.001
How violent the scene was	556.390	1	38	< 0.001
The decade scene occurred	0.040	1	38	0.842
Proportion of people in the image who were law enforcement or military	24.545	1	38	< 0.001
Proportion of people in the image who were perceived to be neither Black nor White	245.912	1	38	< 0.001

Proportion of people in the image who were perceived to be White	0.709	1	38	0.405
Proportion of people in the image who were perceived to be Black	5.026	1	38	0.031
Perceived valence of emotional expression of people in the image	204.850	1	38	< 0.001
Perceived age of people in the image	0.005	1	38	0.945

Results found that the neutral images were rated as being equally familiar, occurred in the same decade, and individuals in the images were viewed as equal in perceived age compared to the images depicting perceived White victim(s). Additionally, the neutral images were rated as being significantly less violent, and the individuals in the neutral images were rated as having more positive emotional expressions than those in the images with perceived White victim(s). However, unlike the images depicting perceived Black victim(s), the images of perceived White victim(s) were viewed as less likely to take place in the United States compared to the neutral images.

Table 6 Descriptive Statistics for Each Condition of Victim Race

Rating Dimension	M_{Black}	SD_{Black}	$M_{Neutral}$	$SD_{Neutral}$	M_{White}	SD_{White}
Proportion of someone in the scene actively being harmed	0.806	0.110	0.982	0.021	0.867	0.125

Proportion of raters who indicated scene was not familiar to them	0.992	0.016	0.984	0.014	0.990	0.017
Likelihood scene occurred in The United States	5.175	0.534	4.864	0.327	4.180	0.682
How violent the scene was	4.620	0.825	1.287	0.071	4.482	0.602
The decade scene occurred	4.432	1.111	4.016	1.225	4.081	0.792
Proportion of aggressors who were law enforcement or military	0.771	0.324	-	-	0.565	0.464
Proportion of victims who were law enforcement or military	0.018	0.022	-	-	0.010	0.014
Proportion of individuals in the scene were law enforcement of military	-	-	0.028	0.027	-	-
Proportion of aggressors who were perceived to be neither Black nor White	0.042	0.059	-	-	0.016	0.034
Proportion of aggressors who were perceived to be White	0.950	0.064	-	-	0.972	0.046
Proportion of aggressors who were perceived to be Black	0.008	0.011	-	-	0.013	0.017
Proportion of victims who were perceived to be neither Black nor White	0.024	0.021	-	-	0.876	0.206
Proportion of victims who were perceived to be White	0.024	0.021	-	-	0.876	0.206

Proportion of victims who were perceived to be Black	0.946	0.058	-	-	0.044	0.010
Aggressor's perceived valence of emotional expression	2.825	0.809	-	-	2.566	0.496
Aggressor's perceived age	2.752	0.435	-	-	2.608	0.680
Victim's perceived valence of emotional expression	2.197	0.505	-	-	2.307	0.455
Victim's perceived age	1.966	0.4748	-	-	2.36200	0.8300
Proportion of people in the scene perceived as neither Black nor White	-	-	0.052	0.044	-	-

Proportion of people in the scene perceived as White	-	-	0.946	0.044	-	-
Proportion of people in the scene perceived as Black	-	-	0.003	0.005	-	-
Average valence of emotional expression of people in the scene	-	-	4.063	0.378	-	-
Average perceived age of people in scene	-	-	2.467	1.049	-	-

Descriptive statistics from the final equated set of images of violent scenes and mundane social interactions.

Conclusion

Complementing Study 1A, in Study 1B, we compiled, piloted, and equated a set of images that depict actions of brutal violence and mundane social interactions. The violent images were comprised of one or more perceived White aggressors attacking one or more victims who were either perceived to be Black Men or

perceived White Men. These images can provide a means to examine how intergroup dynamics (i.e., relative preferences to ingroup members) may influence neural activity in brain regions implicated in social evaluation, mentalizing/ individuation, and salience/ aversion.

Chapter 4

STUDY 2: PROCESSING OF AFFILIATIVE INTERGROUP INTERACTIONS

Existing work that has examined the neural correlates of intergroup bias has generally been reliant on the use of face stimuli (e.g., Barth et al., 2020; Cloutier, Freeman, et al., 2014; Dang et al., 2019; Mattan, Kubota, Dang, et al., 2018; Mattan, Kubota, Li, et al., 2018) which while informative may lack the ecological validity of everyday social interactions. To this end, Study 2 leverages depictions of more complex social interactions to examine how perceivers may differentially process depictions of intergroup affiliation. This study used images of romantic couples where both partners are perceived as Black, both perceived as White and where the masculine presenting partner is perceived as perceived Black and the feminine presenting partner is perceived as White. In the context of the very limited work that involved perceivers viewing depictions of Interracial couples, White perceivers tended to show greater activity in the AI that was thought to reflect negative appraisals of Interracial couples relative to perceived White ones (Skinner & Hudac, 2017). Considering these findings, if perceivers show negative bias toward couples that are not solely comprised of ingroup members, it is possible that perceivers may also show preference toward affiliative depictions of ingroup members. That is, if ingroup preferences are also present, individuals may experience heightened reward when

viewing racially homogeneous couples, particularly of their own ingroup, and may show aversion when exposed to depictions of Interracial couples. These responses may be driven by a combination of positive ingroup valuation. As such, in Study 2, we examined brain regions implicated in social evaluation (OFC, VS) salience detection (AI, amygdala) when participants viewed images of couples comprised of two same-race men or women (i.e., a perceived Black woman and a perceived Black man, perceived White woman and a perceived White man), or ones comprised of a perceived Black man and a perceived White woman. Additionally, we examined brain regions implicated in bias regulation (DLPFC, VLPFC, DACC) in an exploratory fashion.

Hypotheses

Our primary hypotheses centered on brain regions supporting social evaluation and reward, as well as aversion/salience monitoring. Additionally, we offered hypotheses for our exploratory analysis of regions implicated in intergroup bias regulation. As such, in the following paragraphs, we highlight the expected results from brain regions supporting these mechanisms.

We had two tentative hypotheses for regions supporting social reward (i.e., the OFC and NACC). First, because of previous work that found both the NACC (Cikara et al., 2011; Hackel et al., 2015, 2020; Stanley et al., 2012; Wang et al., 2019) and OFC (Cassidy & Krendl, 2016; Firat et al., 2017; Tamir & Hughes, 2018; Willis et al.,

2010) have reflected more positive evaluations toward ingroup members, we predicted that participants would show decreased activity in both the OFC and NACC to depictions of perceived non-White couples relative to perceived White couples. Second, because the sample of participants is comprised of self-identified White adults from the general public, we tentatively predicted that participants would show greater reward responses to perceived Black relative to Interracial couples, considering previous work that found the OFC and NACC may be sensitive to outgroup members particularly when conflict or intergroup biases are more salient (Cikara et al., 2011; Cikara & Fiske, 2012). Because of this, it is possible that both the OFC and NACC may show greater activity toward Black couples relative to Interracial couples based on previous findings showing a sample of perceived White American participants evaluating perceived Black couples more positively than Interracial couples (Skinner & Hudac, 2017).

We offered diverging hypotheses for the regions that support aversion and salience monitoring. While Interracial dating and marriage are becoming increasingly commonplace, with nearly 20% of married couples identifying as Interracial (Pandey, 2022; K. Parker & Barroso, 2021), Interracial relationships are still relatively less common than same-race relationships and may be more novel to some perceivers. As such, we first predicted that because of the nature of Interracial relationships being more novel to some perceivers, it is possible that both the AI and the amygdalae would respond preferentially to perceived non-White couples relative to perceived White couples, as our sample of White American participants may have less expertise

with perceived non-White couples. This, in turn, would suggest that participants would view the outgroup couples as more salient due to their relative novelty compared to ingroup couples. Because the amygdala has been thought to support salience detection in support of social learning (Molapour et al., 2015) rather than being implicated directly in supporting negative intergroup evaluations (Chekroud et al., 2014), we did not predict that differences in amygdala activity would reflect evaluative biases toward either perceived Black or Interracial couples. However, because the AI has been shown to be associated with aversion (Borg et al., 2008) and negative evaluations of Interracial couples (Skinner & Hudac, 2017), we did predict to observe greater AI activity toward Interracial couples relative to the perceived White couples.

The DACC and DLPFC are often posited to support the regulation of intergroup bias. However, existing neuroscientific work suggests that these regulatory mechanisms are used when people are motivated by a strong intrinsic desire to counter existing biases (Amodio et al., 2004, 2006, 2008). Given previous work that found more negative appraisals of depictions of Interracial couples relative to perceived White couples, it is possible that participants may detect conflict between their desires to be egalitarian and their appraisals of Interracial couples. Therefore, as an exploratory analysis of these regions, we predicted that participants may show greater activity in the VLPFC, DLPFC, and the DACC toward Interracial couples relative to perceived White couples, as they may be more motivated to regulate less positive evaluations toward the Interracial couples relative to the perceived White couples. As

such, the confirmatory analysis for Study 2 centered on the OFC, the NACC, amygdalae, and AI, while the exploratory analyses focused on the DACC, DLPFC, and VLPFC.

Methods

Participants

All potential participants were screened prior to their experimental sessions to determine eligibility. Eligible participants self-identified as White and non-Hispanic, were right-handed without a history of severe head injury or fainting, and otherwise met the University of Delaware's Center for Biomedical and Brain Imaging's (CBBI) safety criteria based on responses to a pre-study eligibility survey. Fifty-two participants were recruited from the University of Delaware and the surrounding community. One participant was excluded from the final analysis after indicating they did not self-identify as White and non-Hispanic on the post-scan demographics measure despite self-identifying on the pre-study eligibility survey. One additional participant was removed from the final analysis for excessive head movement, leaving a final sample of 49 participants (39 Women, 9 Men, 1 Genderfluid; M_{age} = 19.143, SD_{age} = 2.336). All participants provided informed consent in accordance with the guidelines set by the Declaration of Helsinki, and the study protocol was approved by the University of Delaware's Institutional Review Board.

Experimental Procedure

Impression Formation Task

Participants then completed two runs of an impression formation task using stimuli depicting smiling couples curated in Study 1A. In this task, participants were instructed to evaluate each image that they saw. Images were presented two times within the study, once per functional run. Each trial of the impression formation task consisted of a centrally presented fixation cross displayed for 500ms, followed by image presentation for 2000ms. Following the presentation of the scene, participants would view a jittered fixation for intervals of 0, 2000, 4500, or 7000ms before the next trial would begin. Prior to the scan, participants were instructed to view and passively form impressions of the images of couples as they appeared on screen. They were tasked with pressing buttons corresponding to their thumbs on two button boxes they were given during the scan to ensure that they were actively attending to the images. Additionally, because their button presses did not differ for each trial, they were told that the impressions they formed were private. Each run consisted of a total of 60 trials, and runs lasted approximately 5 minutes each.

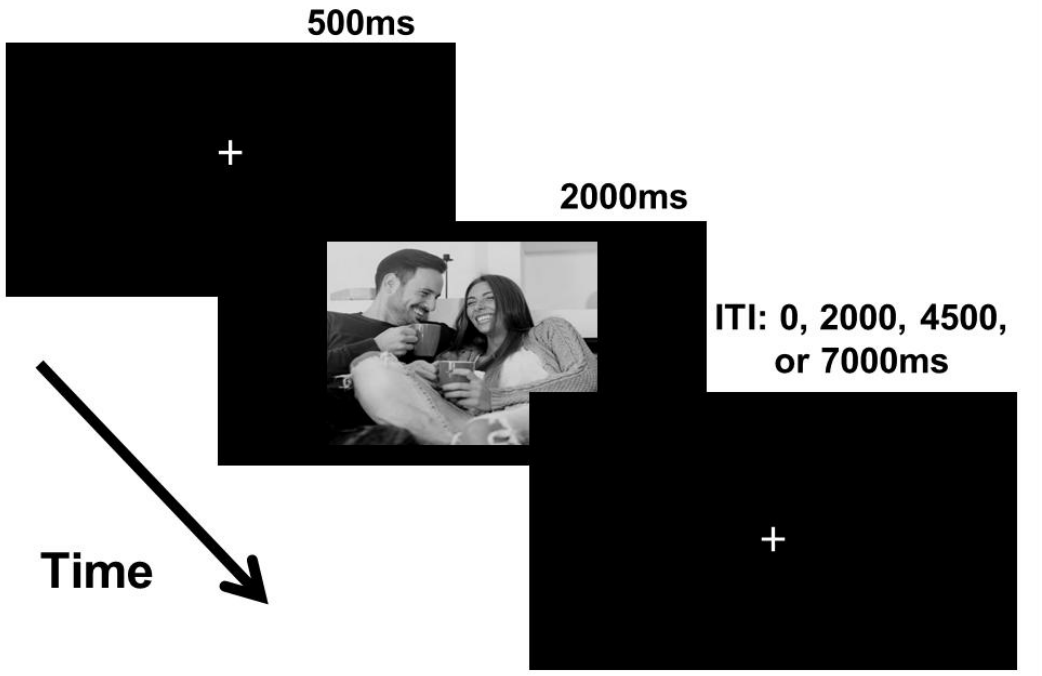

Figure 6. Schematic depicting trial structure for the impression formation task. In each trial, participants would view a centrally presented fixation cross for 500ms, then would view and passively form an impression of the image during its 2000ms presentation window. A jittered intertrial fixation cross (ITI) would then be presented for 0, 2000, 4500, or 7000 ms.

Stimuli

The images used in Study 2 were the images of couples that were piloted and normed during Study 1A of this dissertation.

(A)

(B)

(C)

Figure 7. Examples of images of perceived Black (Panel A), Interracial (Panel B), and perceived White (Panel C) couples.

fMRI Data Acquisition

MRI data were acquired on a 3T Siemens Magneton Prisma MRI scanner with a 64-channel head coil (Siemens Medical, Erlangen, Germany) at the University of Delaware Center for Biomedical and Brain Imaging. For each subject, functional images were collected in 2 functional runs of 120 TRs each, interleaved in 38 axial

slices (3.6mm thick) with a T2*-weighted echo-planar imaging (EPI) sequence: MB acceleration factor = 4, repetition time (TR) = 2500 ms, echo time (TE) = 27 ms, flip angle (FA) = 70 degrees, voxel size = 2.6 mm * 2.6 mm * 3.6 mm, field of view (FOV) = 230 mm * 230 mm. High-resolution structural images were acquired in the sagittal plane using a T1-weighted 3D Turbo Field Echo (TFE/MP-Rage) anatomical scan (TR = 2080ms, echo time = 4.64ms, FOV = 210 × 210 mm, slice thickness = 0.70 mm, voxel size = 0.70 × 0.70 × 0.70 mm).

Slices were aligned to the anterior commissure—posterior commissure axis of each participant. Each run lasted approximately 5 minutes. Timing sequences for each run were created using optseq2 (Greve, 2002). Each sequence was designed to be distinct and ensured that a stimuli condition was not presented more than three consecutive times. Sequence order was counterbalanced across participants. Functional imaging data were pre-processed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) and SPM12W (https://github.com/wagner-lab/spm12w) to remove noise and to correct for motion artifacts. The images were then transformed into a standard anatomical space (3mm isotropic voxels) based on the ICBM 152 template (MNI, Montreal Neurological Institute). Normalized data were then spatially smoothed using an 8mm full width at half maximum (FWHM) Gaussian Kernal to increase the signal-to-noise ratio and to reduce the impact of anatomical variability not corrected for by stereotaxic normalization. GLMs were constructed for each participant to examine condition-specific brain activity as a function of the couple's racial composition (i.e., perceived Black couples, perceived

White couples, Interracial couples). Time and dispersion derivatives were also modeled for each of the couple conditions.

Data Analysis

Region of interest (ROI) fMRI data were analyzed using the LME4 package (Bates et al., 2015) in the R programming language (R Core Team, 2017). Analyses consisted of a within-subjects design and included an intercept for each subject. The within-subjects independent factor was couple racial composition. Three models were run to compare across all three levels of the within-subjects variable of couple race. The first model compared perceived White couples to perceived Black and Interracial couples (contrast code: -0.33 = perceived Black couples, -0.33 = Interracial couples, 0.66 = perceived White couples) and additionally compared perceived Black couples directly to Interracial couples (0.5 = Interracial couples, -0.5 = perceived Black couples). The second model compared Interracial couples to perceived Black and perceived White couples (-0.33 = perceived Black couples, -0.33 = perceived White couples, 0.66 = Interracial couples) and additionally compared perceived White couples to the perceived Black couples (0.5 perceived White couples, -0.5 = perceived Black couples). The third model compared perceived Black couples to perceived White and Interracial couples (-0.33 = Interracial couples, -0.33 = perceived White couples, 0.66 = perceived Black couples) and additionally compared perceived White to Interracial couples (0.5 = perceived White couples, -0.5 = Interracial couples).

ROI Analysis

Region of interest (ROI) analyses focused on regions that have been implicated in social evaluations of others, such as the OFC and the NACC. We also examined ROIs implicated in the salience detection and negative evaluative responses, such as the amygdala and AI, and those implicated in the regulation of intergroup bias, such as the DACC, DLPFC, and VLPFC.

Bold activity in bilateral OFC was taken from a previous investigation on social evaluations, left OFC ($MNI_{x,y,z}$ = -9, 40, -15; 8 mm sphere), right OFC ($MNI_{x,y,z}$ = 9, 40, -15; 8 mm sphere; Cloutier, Heatherton, et al., 2008). Bilateral NACC and amygdalae ROIs were modeled based on a previous investigation that examined race and status-based social evaluations left NACC ($MNI_{x,y,z}$ = -9, 8, -8,; 4 mm sphere), right NACC ($MNI_{x,y,z}$ = 9, 14, -6; 4 mm sphere; Mattan, Kubota, Dang, et al., 2018) left amygdala ($MNI_{x,y,z}$ = -24, -3, -21; 4 mm sphere), right amygdala ($MNI_{x,y,z}$ = 24, -3, -21; 4 mm sphere; (Mattan, Kubota, Dang, et al., 2018). Bilateral AI ROIs were modeled based on existing meta-analyses that examined the AI and DACC's role in salience detection and emotional regulation, left AI ($MNI_{x,y,z}$ = -42, 16, 0, 8 mm sphere), right AI ($MNI_{x,y,z}$ = 36, 18, 0; 8 mm sphere; Adolfi et al., 2017), left DACC ($MNI_{x,y,z}$ = -4, 20, 38; 8 mm sphere), right DACC ($MNI_{x,y,z}$ = 6, 26, 38; 8 mm sphere; Zhang & Peng, 2023).

We examined the beta weights (generated from the GLMs) of all ROIs as a function of couple racial composition. For all ROI analyses, we were only able to model a random intercept for each participant. To account for multiple comparison concerns across confirmatory ROIs (n = 4), a Bonferroni correction was applied to each confirmatory ROI with an adjusted $\alpha = 0.0125$. While all significant results for both confirmatory and exploratory regions are reported, effects from confirmatory ROIs that do not survive the Bonferroni correction should be interpreted cautiously.

Whole-Brain Analysis

Complementary to the ROI analyses, we conducted an exploratory whole-brain analysis on our main contrasts of interest: White couples > Black couples, White couples > Interracial couples, and Black couples > Interracial couples. We used an uncorrected voxel level of $p < 0.001$ and a cluster extent threshold of 15 voxels. All whole-brain analyses should be treated as exploratory and interpreted in the context of the ROI results.

Results

Button Press Response Times

To ensure that participants were not actively attending to one stimulus condition more rapidly than another, we analyzed button press response times for each couple's condition using the LME4 package in R (Bates et al., 2015) across three

models. The first model compared perceived White couples to perceived Black and Interracial couples (contrast code: -0.33 = perceived Black couples, -0.33 = Interracial couples, 0.66 = perceived White couples) and additionally compared perceived Black couples directly to Interracial couples (0.5 = Interracial couples, -0.5 = perceived Black couples). The second model compared Interracial couples to perceived Black and perceived White couples (-0.33 = perceived Black couples, -0.33 = perceived White couples, 0.66 = Interracial couples) and additionally compared perceived White couples to the perceived Black couples (0.5 perceived White couples, -0.5 = perceived Black couples). The third model compared perceived Black couples to perceived White and Interracial couples (-0.33 = Interracial couples, -0.33 = perceived White couples, 0.66 = perceived Black couples) and additionally compared perceived White to Interracial couples (0.5 = perceived White couples, -0.5 = Interracial couples).

All response times were log-transformed prior to analysis, and all trials with no button responses were removed prior to analysis. Results revealed no significant differences between couple conditions, $ps > 0.367$ (M_{Black} = 1155.498ms, SD_{Black} = 414.881ms, $M_{Interracial}$ = 1154.125ms, $SD_{Interracial}$ = 406.510ms, M_{White} = 1175.733ms, SD_{White} = 423.181ms).

Left OFC

Results revealed significantly greater BOLD activity in the left OFC to Interracial couples relative to non-Interracial couples, $b = 0.048$, $SE = 0.016$, $CI_{95\%} = [0.016,0.080]$, $t(96) = 2.948$, $p = 0.004$. This main effect was further qualified by greater activity toward Interracial couples relative to perceived Black couples, $b = -0.052$, $SE = 0.019$, $CI_{95\%} = [-0.089, -0.016]$, $t(96) = -2.822$, $p = 0.006$ (see Figure 8). There was additionally diminished activity toward perceived White couples relative to Interracial couples, $b = -0.042$, $SE = 0.019$, $CI_{95\%} = [-0.079, -0.006]$, $t(96) = -2.283$, $p = 0.025$, though this effect did not survive Bonferroni correction. All other main effects were not significant, $p > 0.055$.

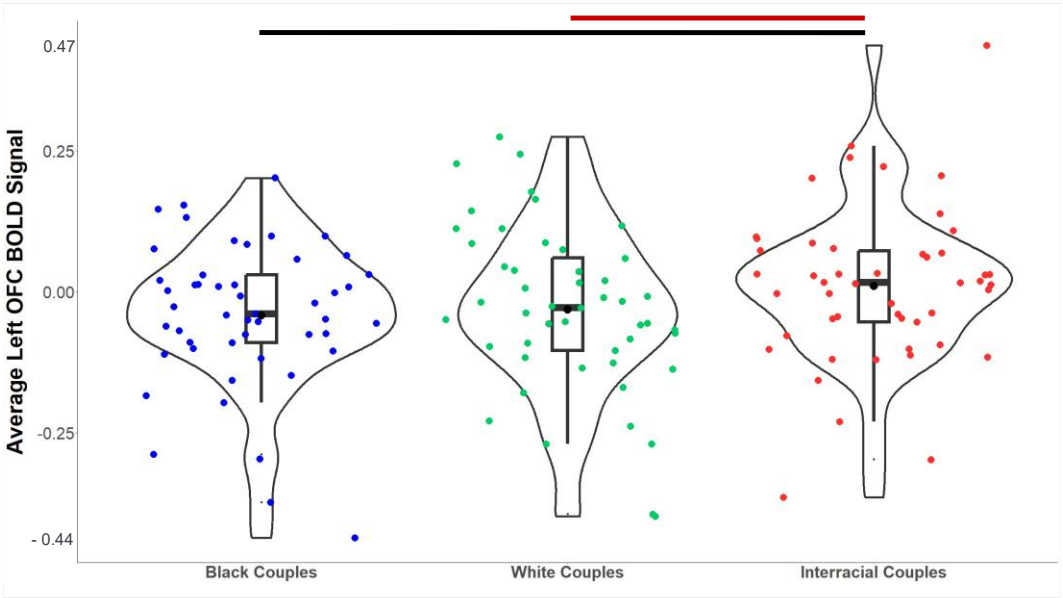

Figure 8. BOLD activity in the Left OFC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples. However, the difference between perceived White couples and Interracial couples did not survive Bonferroni correction.

Right OFC

Results in the right OFC revealed greater activity toward Interracial couples relative to non-Interracial couples, $b = 0.046$, $SE = 0.019$, $CI_{95\%} = [0.010, 0.081]$, $t(96) = 2.501$, $p = 0.014$, though this effect did not survive Bonferroni correction. Additionally, there was significantly greater activity to Interracial relative to perceived Black couples, $b = -0.010$, $SE = 0.19$, $CI_{95\%} = [-0.094, -0.012]$, $t(96) = -2.556$, $p = 0.012$ (see Figure 9). All other effects were not significant, $p > 0.057$.

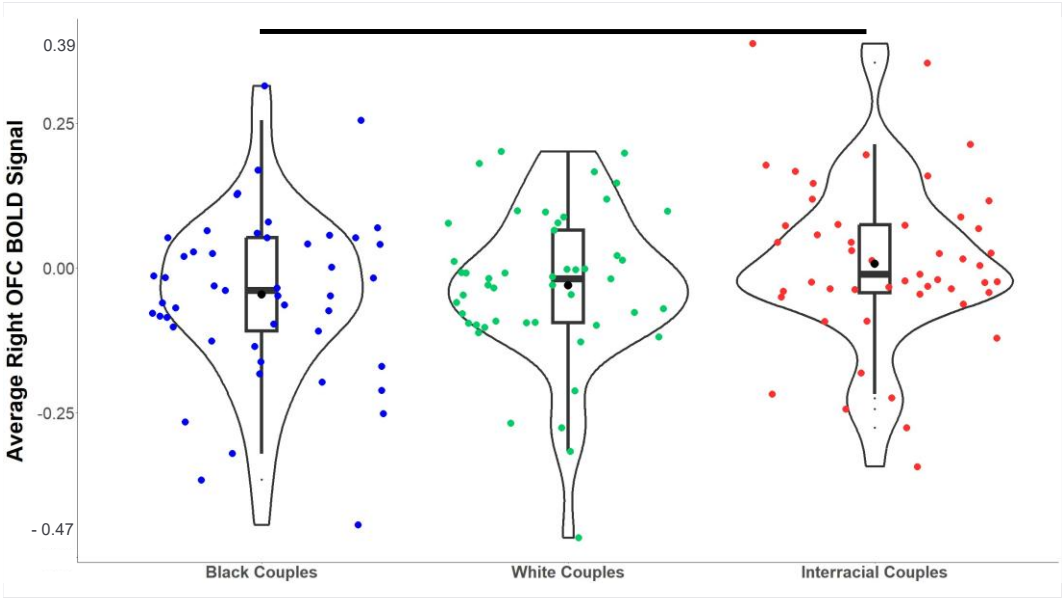

Figure 9. BOLD activity in the Right OFC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to perceived Black couples.

Left NACC

Results revealed greater activity toward Interracial couples relative to either perceived Black or perceived White couples, b = 0.041, SE = 0.015, $CI_{95\%}$ = [0.013, 0.070], $t(96)$ = 2.86, p = 0.005. Further, greater activity was observed toward Interracial couples relative to perceived Black couples, b = -0.0514, SE = 0.0165, $CI_{95\%}$ = [-0.084, -0.019], $t(96)$ = -3.11, p = 0.003 (see Figure 10). Additionally, there

was greater activity to perceived non-Black relative to perceived Black couples, though this did not survive Bonferroni correction, b = -0.036, SE = 0.015, $CI_{95\%}$ = [-0.065, -0.008], $t(96)$ = -2.52, p = 0.014. All other effects were not significant, 0.067.

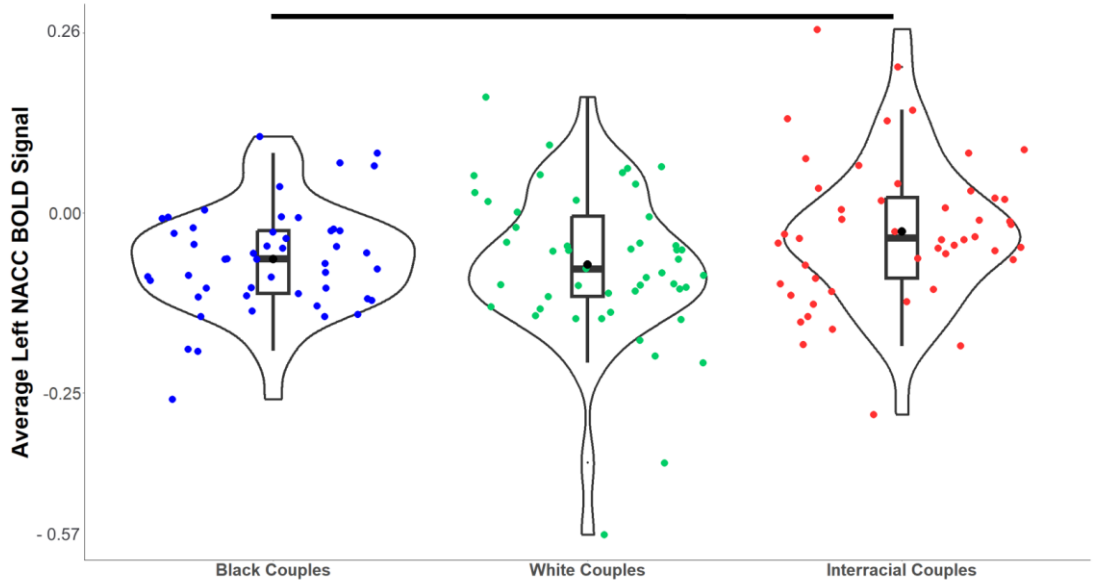

Figure 10. BOLD activity in the Left NACC as a function of couple race. Blue dots represent depictions of perceived Black couples, Green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to perceived Black couples.

Right NACC

Results in the right NACC revealed greater activity toward Interracial couples relative to either perceived Black or perceived White couples, b = 0.043, SE = 0.014,

$CI_{95\%}$ = [0.015, 0.070], $t(96)$ = 3.06, p = 0.003. Further, greater activity was observed toward Interracial couples relative to perceived White couples, b = -0.046, SE = 0.016, $CI_{95\%}$ = [-0.077, -0.0148], $t(96)$ = 2.89, p = 0.005 (see Figure 11). Additionally, there was greater activity to Interracial relative to perceived Black couples, though this did not survive Bonferroni correction, b = -0.0386, SE = 0.0159 $CI_{95\%}$ = [-0.0698, -0.00729], $t(96)$ = -2.42, p = 0.0175. All other effects were not significant, $p > 0.056$.

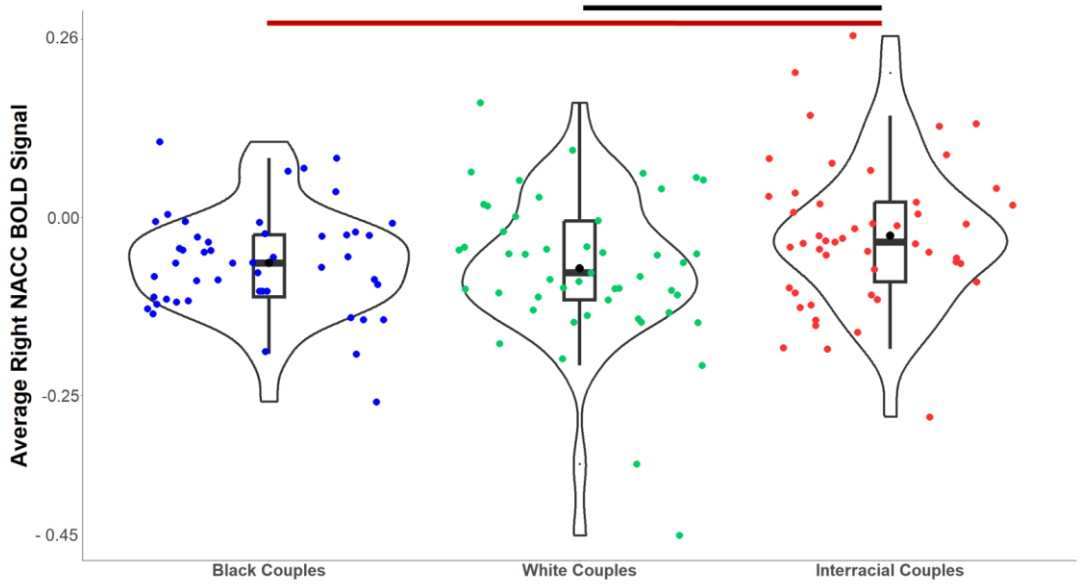

Figure 11. BOLD activity in the Right NACC as a function of couple race. Blue dots represent depictions of perceived Black couples, Green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to both

perceived Black and perceived White couples. However, the difference between perceived Black couples and Interracial couples did not survive Bonferroni correction.

Left Amygdala

Results revealed significantly greater left amygdala activity toward Interracial couples relative to non-Interracial couples, $b = 0.044$, $SE = 0.014$, $CI_{95\%} = [0.018, 0.070]$, $t(96) = 3.27$, $p = 0.001$. This effect was qualified by significantly greater activity in the left amygdala to Interracial couples relative to both perceived White $b = -0.042$, $SE = 0.015$, $CI_{95\%} = [-0.072, -0.012]$, $t(96) = -2.74$, $p = 0.007$, and perceived Black couples $b = -0.045$, $SE = 0.015$, $CI_{95\%} = [-0.075, -0.015]$, $t(96) = -2.93$, $p = 0.004$ (see Figure 12). All other effects were not significant, $p > 0.074$.

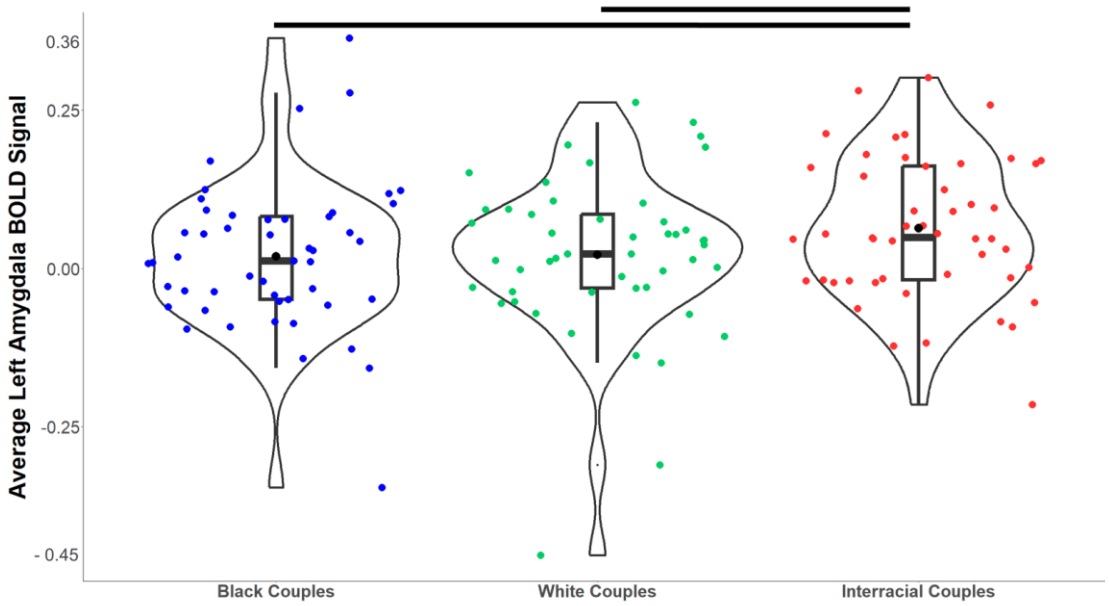

Figure 12. BOLD activity in the Left amygdala as a function of couple race. Blue dots represent depictions of perceived Black couples, Green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The

y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples.

Right Amygdala

Results revealed greater right amygdala activity toward perceived non-Black couples, b = -0.032, SE = 0.015, $CI_{95\%}$ = [-0.062, -0.002], $t(96)$ = -2.06, p = 0.042, though this effect did not survive Bonferroni correction. All other effects were not significant, $p > 0.070$.

Left AI

Results in the left AI found greater activity toward Interracial couples relative to non-Interracial couples b = 0.117, SE = 0.035, $CI_{95\%}$ = [0.048, 0.186], $t(96)$ = 3.33, p = 0.001. This effect was qualified by greater activity to Interracial couples relative to perceived White couples, b = -0.104, SE = 0.040, $CI_{95\%}$ = [-0.182, -0.025], $t(96)$ = -2.58, p = 0.011, as well relative to perceived Black couples, b = -0.128, SE = 0.0401, $CI_{95\%}$ = [-0.207, -0.050], $t(96)$ = -3.19, p = 0.002 (see Figure 13). There was also significantly greater activity to perceived non-Black couples relative to perceived Black couples. However, this effect did not survive Bonferroni correction, b = -0.077,

SE = 0.035, $CI_{95\%}$ = [-0.146, -0.008], *t*(96) = -2.20, *p* = 0.030. All other effects were not significant, *p* > 0.259.

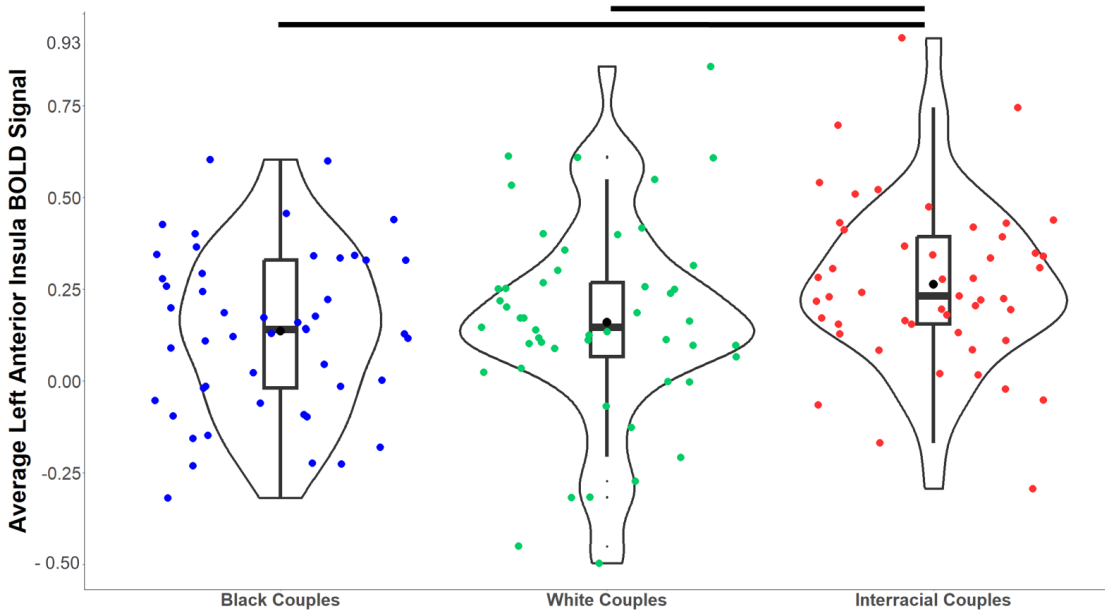

Figure 13. BOLD activity in the Left AI as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples.

Right AI

Results found greater activity to Interracial couples relative to non-Interracial couples, *b* = 0.082, *SE* = 0.024 $CI_{95\%}$ = [0.034, 0.129], *t*(96) = 3.38, *p* = 0.001. This effect was qualified by greater activity to Interracial relative to perceived Black

couples, b = -0.102, SE = 0.028, $CI_{95\%}$ = [-0.156, -0.048], $t(96)$ = -3.68, p<0.001 (see Figure 14). Additionally, there was greater activity to Interracial relative to perceived White couples, though this did not survive Bonferroni correction, b = -0.060, SE = 0.028, $CI95\%$ = [-0.115, -0.006], $t(96)$ = -2.17, p = 0.0322. There was also greater activity to perceived non-Black relative to perceived Black couples, b = -0.0724, SE = 0.0242, $CI_{95\%}$ = [-0.12, -0.025], $t(96)$ = -2.99, p = 0.004. All other effects were not significant, p > 0.136.

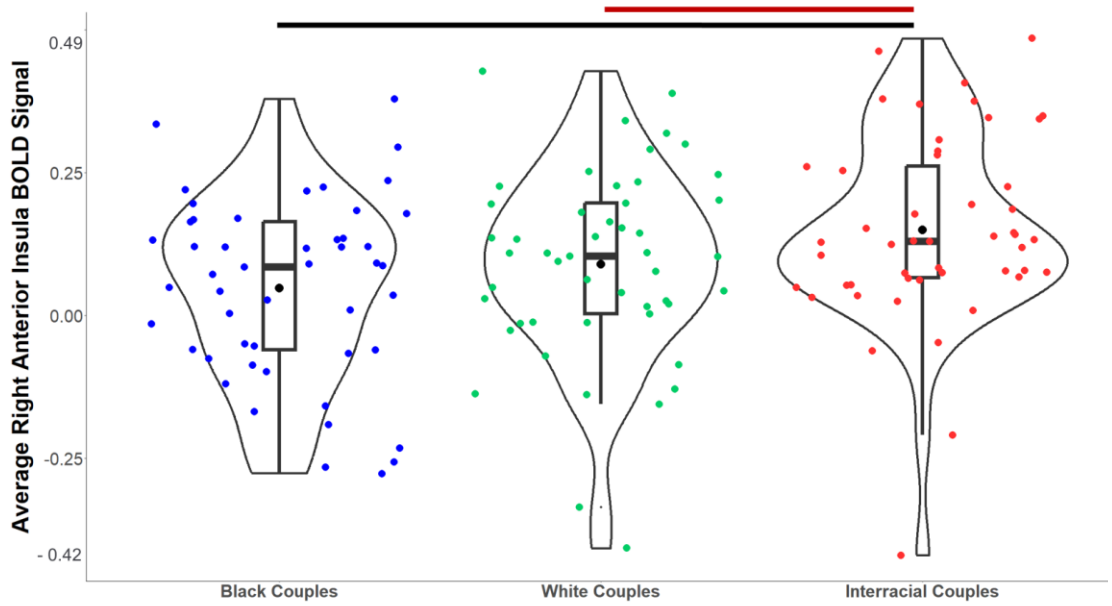

Figure 14. BOLD activity in the Right AI as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas

red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples. However, the difference between perceived White couples and Interracial couples did not survive Bonferroni correction.

Exploratory ROI

Left DACC

Results in the left DACC revealed greater activity toward Interracial relative to non-Interracial couples, $b = 0.082$, $SE = 0.027$, $CI_{95\%} = [0.029, 0.136]$, $t(96) = 3.04$, $p = 0.003$. This effect was qualified by greater activity to Interracial relative to perceived White couples, $b = -0.081$, $SE = 0.031$, $CI_{95\%} = [-0.142, -0.021]$, $t(96) = -2.63$, $p = 0.01$ and greater activity to Interracial couples relative to perceived Black couples, $b = -0.082$, $SE = 0.031$, $CI_{95\%} = [-0.142, -0.021]$, $t(96) = -2.64$, $p = 0.001$ (see Figure 15). All other effects were not significant, $p > 0.128$.

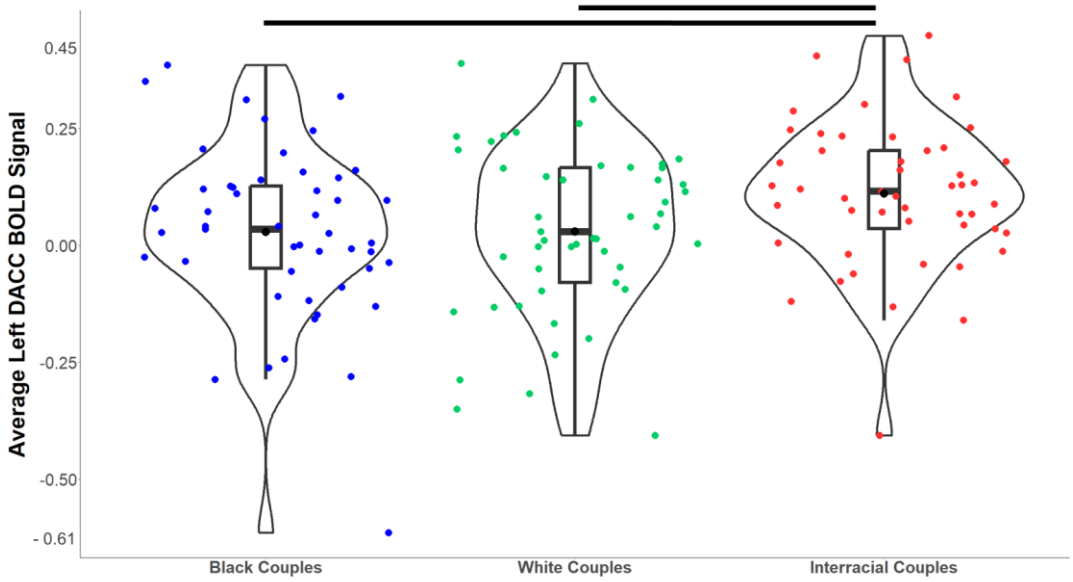

Figure 15. BOLD activity in the Left DACC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples.

Right DACC

Results in the right DACC found greater activity toward Interracial couples relative to non-Interracial couples, $b = 0.073$, $SE = 0.029$, $CI_{95\%} = [0.016, 0.131]$, $t(96) = 2.49$, $p = 0.014$. This effect was driven primarily by greater activity to Interracial relative to perceived White couples, $b = -0.079$, $SE = 0.034$, $CI_{95\%} = [-0.145, -0.013]$, $t(96) = -2.34$, $p = 0.021$, though neither effect survived Bonferroni correction (see Figure 16). All other effects were not significant, $p > 0.051$.

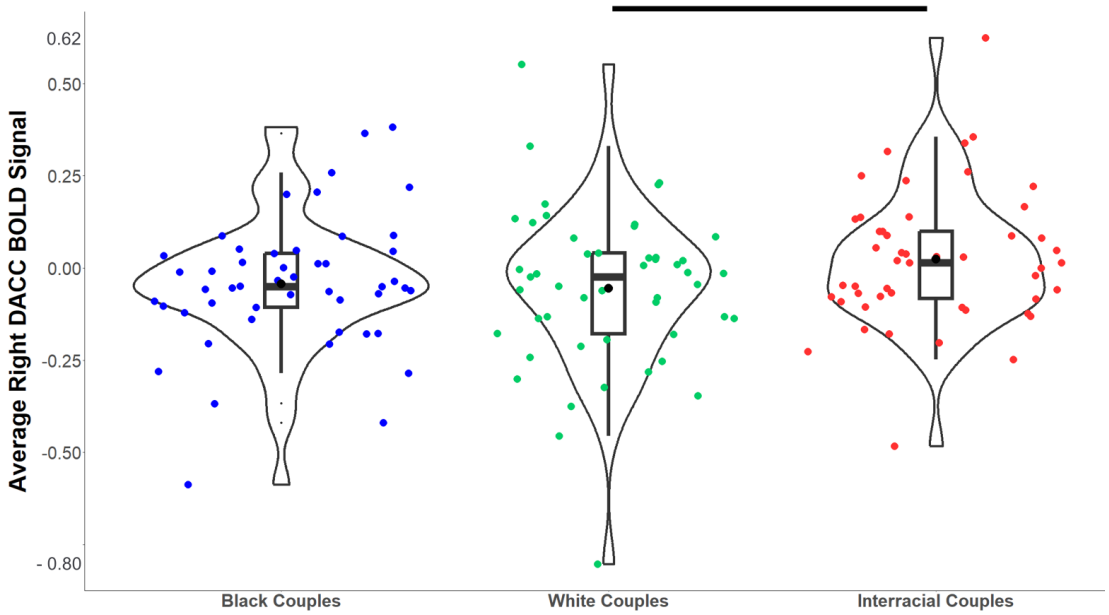

Figure 16. BOLD activity in the Right DACC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to Interracial couples relative to perceived White couples.

Left DLPFC

Results in the left DLPFC found greater activity to Interracial relative to non-Interracial couples, $b = 0.124$, $SE = 0.044$, $CI_{95\%} = [0.038, 0.21]$, $t(96) = 2.82$, $p = 0.006$. This effect was qualified by greater activity to Interracial couples relative to perceived White couples, $b = -0.132$, $SE = 0.050$, $CI_{95\%} = [-0.23, -0.033]$, $t(96) = -2.62$, $p = 0.0102$ (see Figure 17). There was also significantly greater activity to

Interracial relative to perceived Black couples, b = -0.114, SE = 0.050, $CI_{95\%}$ = [-0.212, -0.016], $t(96)$ = -2.27, p = 0.026. All other effects were not significant, p > 0.089.

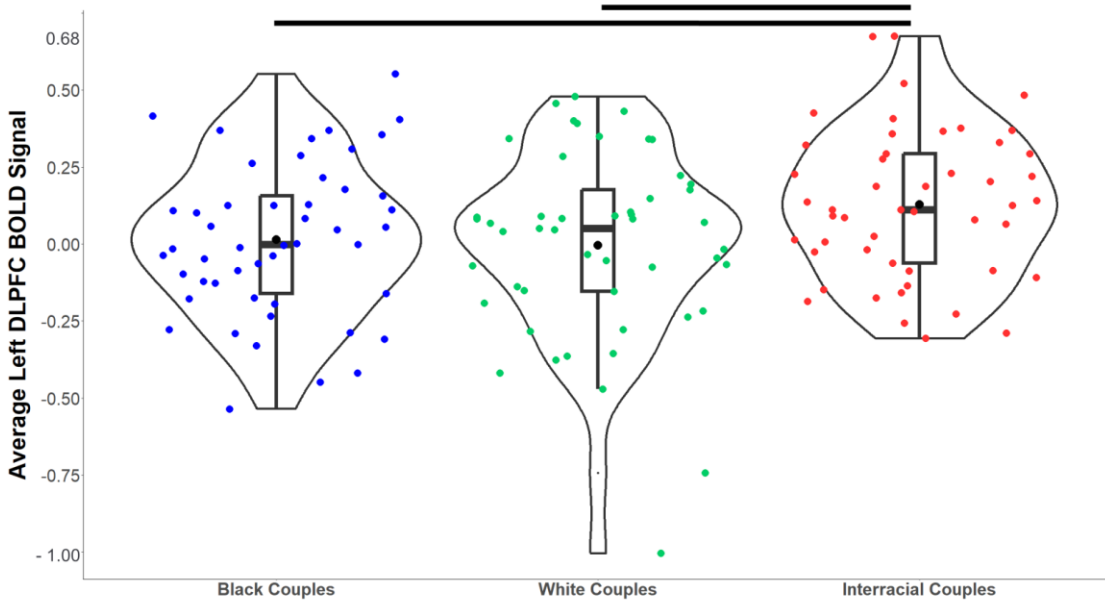

Figure 17. BOLD activity in the Left DLPFC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples.

Right DLPFC

Results revealed greater right DLPFC activity to Interracial relative to non-Interracial couples, b = 0.076, SE = 0.036, $CI_{95\%}$ = [0.006, 0.145], $t(96)$ = 2.13, p = 0.036, as well as greater activity to Interracial relative to perceived White couples, b =

-0.083, SE = 0.041, $CI_{95\%}$ = [-0.162, -0.003], $t(96)$ = -2.04, p = 0.044 (see Figure 18).

All other effects were not significant, $p > 0.102$.

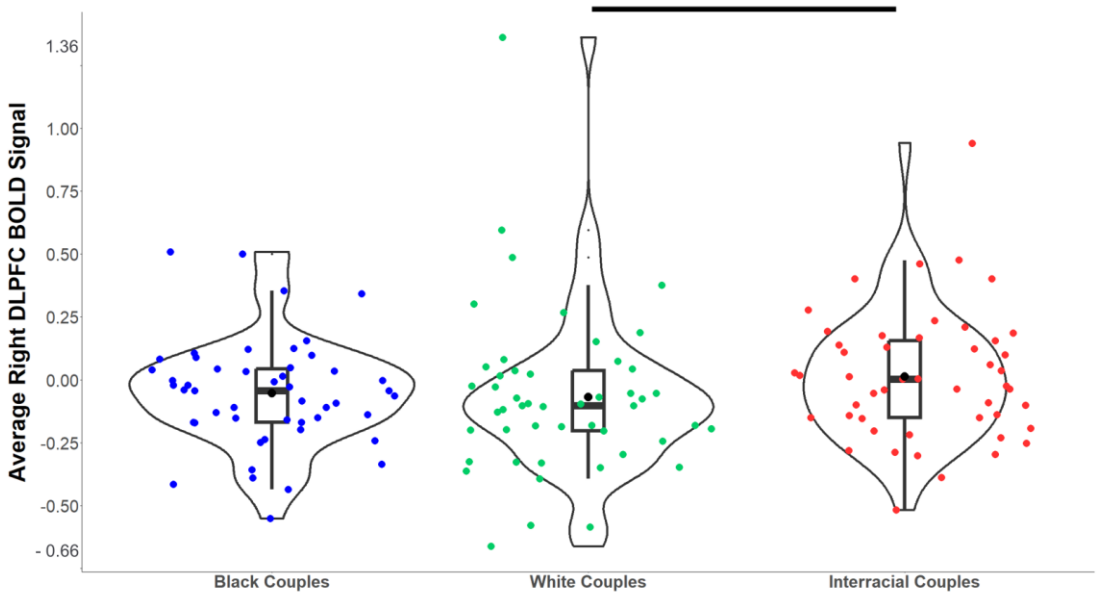

Figure 18. BOLD activity in the Right DLPFC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to Interracial couples relative to perceived White couples.

Left VLPFC

Results in the left VLPFC found greater BOLD activity to Interracial relative to non-Interracial couples, b = 0.087, SE = 0.029, $CI_{95\%}$ = [0.023, 0.144], $t(96)$ = 2.98, p = 0.004. This effect was driven by greater activity to Interracial couples relative to

perceived Black couples, b = -0.091, SE = 0.033, $CI_{95\%}$ = [-0.156, -0.026], $t(96)$ = -2.73, p = 0.007 (see Figure 19). Additionally, there was greater BOLD activity to Interracial couples relative to perceived White couples, b = -0.081, SE = 0.033, $CI_{95\%}$ = [-0.146, -0.015], $t(96)$ = -2.42, p = 0.017. All other effects were not significant, p > 0.082.

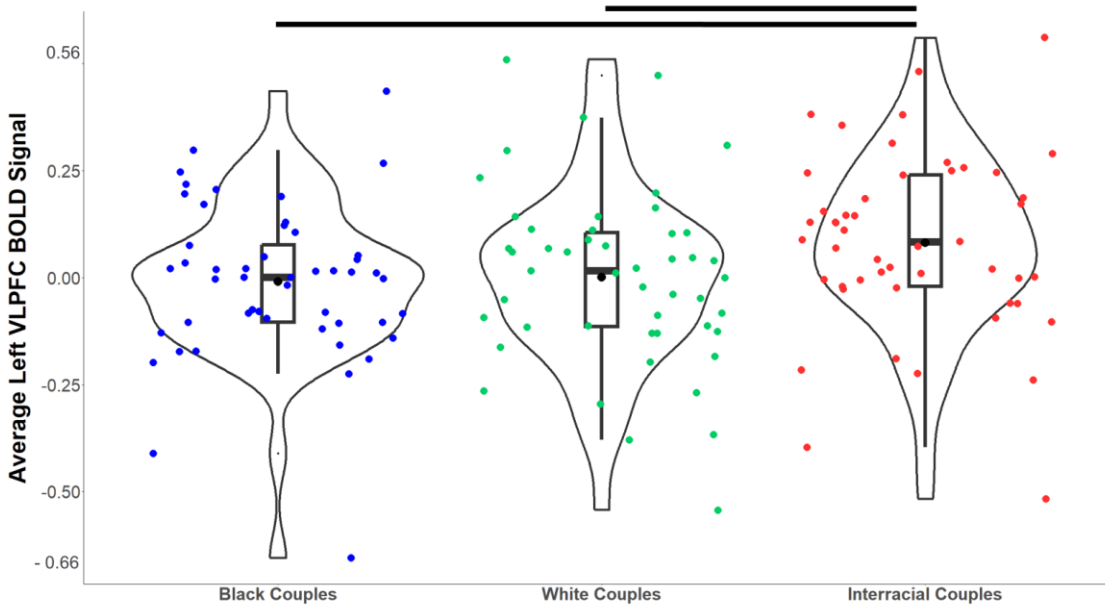

Figure 19. BOLD activity in the Left VLPFC as a function of couple race. Blue dots represent depictions of perceived Black couples, green dots represent depictions of perceived White couples, and red dots represent depictions of Interracial couples. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to Interracial couples relative to both perceived Black and perceived White couples. However, the difference between perceived White couples and Interracial couples did not survive Bonferroni correction.

Right VLPFC

There were no significant effects in the right VLPFC, $p > 0.154$.

Whole Brain

Table 7 Whole-brain bold signal as a function of Couple Race.

					MNI Coordinates		
Contrast	**Brain Region**	**BA**	***K***	***t-value***	***x***	***y***	***z***
Black couples > Interracial couples	L.Cuneus	19	29	3.93	-18	-93	21
Black couples > White couples	NA						
Interracial couples > Black couples	Culmen		5205	9.23	-3	-45	0
	L.Insula			5.39	-42	3	-3
	R. Insula			4.47	36	12	0
	R. Caudate			4.12	15	-9	18
	L.Amygdala			3.32	-24	-6	-21
	L.Postcentral Gyrus		2371	6.91	-18	-24	72
	R.Medial Frontal Gyrus	6		6.53	6	-18	69
	VMPFC	10	86	4.89	-3	42	-9
	DMPFC		26	4.8	0	42	45
	L.Inferior Frontal Gyrus	47	20	4.39	-30	33	-9
	L.Middle Frontal Gyrus		41	4.25	-27	54	15
	R.Precentral Gyrus		29	4.03	63	-9	24
	R.Middle Frontal Gyrus		50	4.01	54	24	27
	R.Inferior Parietal Lobule		15	3.94	66	-39	24
	Precuneus		37	3.59	0	-60	30

Interracial couples > White couples	L.Middle Temporal Gyrus		528	5.64	-45	3	-27
	L.Temporal Pole	28		4.53	-24	9	-24
	L.Middle Temporal Gyrus		31	5.10	-63	-9	6
	L.Insula	13	45	4.52	-39	6	-3
	L.Middle Occipital Gyrus	39	16	4.33	-51	-78	6
	R.Culmen		25	4.25	15	-30	18
	L.Superior Temporal Gyrus		61	4.20	-36	-27	6
	R.Middle Occipital Gyrus	37	22	3.91	48	-78	-3
	L.Calcarine		15	3.77	0	-93	-3
White couples > Black couples	NA						
White couples > Interracial couples	NA						

Discussion

The present study highlighted how White perceivers may process depictions of affiliative social interactions involving couples comprised of either same-race partners or those comprised of a perceived Black man and a perceived White woman. The results indicated that participants found depictions of Interracial couples to be more rewarding than depictions of same-race couples. Additionally, Study 2 found increased activity in the AI to the images of Interracial couples, partially supporting previous work examining social evaluations of cross-race dating. However, the patterns of activity in the amygdala mirror those in the AI, which may provide

additional context to expand upon in this previous work. Additionally, the exploratory results found that there was greater activity in regions supporting bias regulation and response monitoring to depictions of Interracial couples relative to same-race couples. We elaborate on these findings in the following paragraphs.

The results from Study 2 provide several novel insights. First, the study suggests that this sample of White American perceivers may have viewed depictions of Interracial couples more positively compared to depictions of perceived Black couples, suggested by greater activity toward Interracial couples relative to perceived Black couples in both the OFC and NACC. Additionally, this sample of participants showed greater activity to Interracial relative to white couples. However, this pattern was only present in the NACC, suggesting that this sample of participants may have evaluated Interracial couples more positively than either perceived Black or perceived White couples, seemingly at odds with previous work (Skinner & Hudac, 2017).

While Study 2 did find greater positive evaluations of Interracial couples, it also replicated Skinner & Hudac's (2017) findings of increased activity in the AI to depictions of Interracial couples. However, there are multiple reasons that the findings from Study 2 may better contextualize those reported by Skinner & Hudac. In their investigation, Skinner and Hudac used EEG, specifically the P300 ERP component, as a proxy for AI activity. Their decision to use the P300 component likely stemmed from work concerning the P300 component that characterized the P300 as tracking evaluative responses, and particularly evaluative responses to negative or aversive stimuli (e.g., Bartholow et al., 2001; Broyd et al., 2012; Crites et al., 1995; Delplanque

et al., 2005; Foti et al., 2009; Hajcak & Olvet, 2008; Ito & Cacioppo, 2000) in some ways drawing a parallel to work that has found the AI to be sensitive to aversive stimuli (Borg et al., 2008; Calder et al., 2007; Heining et al., 2003; Ying et al., 2018). Skinner & Hudac ultimately interpreted their findings of increased P300 amplitudes toward depictions of Interracial couples as being indicative of aversion (i.e., disgust). However, both the AI and the P300 also support motivational salience. Within intergroup literature, the P300 has been found to be sensitive to social categories that are salient to perceivers (Bartholow et al., 2006; Correll et al., 2006; Duval et al., 2013; Ito & Urland, 2003, 2005; Kubota & Ito, 2007; Kutas et al., 1977) similar to how the AI is thought to support salience detection (Menon & Uddin, 2010; Uddin, 2015). Taken as a whole, this seems to suggest that in both Study 2 and Skinner & Hudac's investigation, the greater activity in the AI to Interracial relative to same-race couples may possibly be attributed to the increased salience of Interracial couples relative to same-race couples. This interpretation also seems more likely when considering the similar patterns of activity to Interracial couples in the amygdala, where increased activity would possibly be tracking salience responses rather than evaluative ones (Chekroud et al., 2014). As such, Study 2 found that Interracial couples were positively evaluated and, at the same time, viewed as more salient than same-race couples, perhaps due to their relative novelty to this sample of White perceivers.

The exploratory ROI analysis in Study 2 additionally found that participants showed greater activity in the DACC, DLPFC, and VLPFC to Interracial couples

relative to either perceived Black or perceived White couples. All three regions have been implicated in executive control and regulatory processes, specifically with respect to the identification of conflicts during responses and adjusting responses accordingly (Carter et al., 1998; Carter & van Veen, 2007; Egner & Hirsch, 2005; MacDonald et al., 2000; Mansouri et al., 2009; Miller & Cohen, 2001; Ridderinkhof et al., 2004; Roberts & Hall, 2008). The patterns of activity in the DACC and DLPFC, in particular, seem to suggest that this sample of perceivers may have actively exerted executive control over their responses in the interest of them being more in line with their own egalitarian views (Amodio, 2014; Amodio et al., 2006, 2008; Ito & Bartholow, 2009; Kubota et al., 2012; Mattan, Wei, et al., 2018). However, it is difficult to say with certainty that this was in line with participants' own internal or external motivation to regulate prejudice, as neither was measured for the entire sample of Study 2. However, this interpretation seems plausible due to the patterns of activity in the NACC and OFC.

Limitations

While Study 2 found several novel findings, there are certainly some limitations to the findings due to the study design and sample. First, the sample consisted of 49 self-identified White college students who were, for the most part, relatively close to the same age as one another. This, at first, makes it difficult to generalize these findings beyond the scope of White undergraduate students. Second, the study did not measure some individual difference measures that could have better

contextualized the ROI results. One example comes from the interpretation of the results in the AI and amygdala. Including an index of Interracial contact may have helped strengthen the salience interpretation of the patterns of activity in these regions, as it would have further clarified whether the differences in activity were influenced by the degree of exposure to depictions of Interracial couples. However, it does seem likely that these regions were tracking salience due to the novelty of depictions of Interracial couples due to the patterns of results in the OFC and NACC. Similarly, the study did not measure either internal or external motivation to regulate prejudice (IMS/EMS; Plant & Devine, 1998) for the entire sample of participants, which also makes interpretation of the regions supporting regulation more difficult. As such, future investigations that examine the processing of affiliative social scenes similar to these should consider measuring these constructs in the interest of better interpreting patterns of results in brain regions supporting salience and regulation. Nonetheless, Study 2 provided novel insights into how perceivers process more complex affiliative social interactions of cross-race individuals. In doing so, it sets a starting point to compare how perceivers process aversive intergroup social behaviors. Study 2 also lays a foundation for future work that examines how differences in explicit worldviews may further shape appraisals of cross-race harmony.

Conclusion

Study 2 explored the neural correlates of social evaluation, salience processing, and bias regulation during the processing of depictions of cross-race couples. In this

study, we found that greater activity in both the OFC and NACC regions supporting positive social evaluations of Interracial relative to both perceived Black and perceived White couples. Furthermore, in Study 2, we found greater activity in the AI and amygdalae to depictions of Interracial relative to either perceived Black or perceived White couples, partially replicating previous work that examined how individuals process depictions of Interracial couples. However, taken with the patterns of results from regions thought to support positive social appraisals, Study 2 expands upon this previous literature, suggesting that individuals may not view Interracial couples as aversive but rather possibly as more salient. Lastly, exploratory ROI and whole-brain analyses revealed that regions supporting bias regulation may also possibly be preferentially implicated in the processing of Interracial couples.

Chapter 5

STUDY 3: PROCESSING OF AVERSIVE INTERGROUP INTERACTIONS

Social neuroscientific work that has examined the processing of aversive intergroup interactions has generally found pro-ingroup biases among perceivers. Previous investigations have found that brain regions supporting aversion have shown preferential activation when a member of their ingroup has been the target of a negative outcome (Bruneau et al., 2012; Cikara et al., 2011). Furthermore, while processing scenes depicting intergroup conflict, it would be expected for perceivers to try to infer the intentions of the individuals in the scene, particularly those who are of greater personal significance to them (Bruneau et al., 2012), or perhaps more relevant to depictions of violence, those who are believed to be treated violently unfairly (Dang et al., 2022; Molenberghs et al., 2015). As such, examining brain regions that support aversion processing and mentalizing may be a valid starting point to better understand how depictions of cross-race violence may be differently processed by a sample of White perceivers.

Therefore, where Study 2 examined the neural mechanisms that support salience detection, mentalizing, and regulation when a sample of White Americans viewed images of smiling couples who varied in racial composition, Study 3 examines brain regions implicated in salience detection, aversion, mentalizing, and regulation while processing depictions of cross-race violence.

Hypotheses

In Study 3, our confirmatory hypotheses centered on brain regions supporting mentalizing, such as the STS and TPJ, and aversion/salience monitoring, such as the AI and amygdala. Mirroring Study 2, our exploratory hypotheses centered on regions supporting bias regulation, such as the DACC, DLPFC, and VLPFC. Additionally, we examined the DMPFC, which is often implicated in mentalizing in an exploratory fashion. As such, in the following paragraphs, we highlight the expected results from brain regions supporting these mechanisms.

With respect to the AI and the amygdala, we also offered two tentative hypotheses. First, we predicted that there would be an increase in neural activity toward violent images relative to the neutral ones, reflecting greater overall salience of the violent images relative to the neutral ones. Second, if there was a difference in negative evaluative responses in the AI, we tentatively predicted participants may have shown more negative evaluative responses to depictions of violence toward perceived White men relative to depictions of violence toward perceived Black men, reflecting a general pro-ingroup evaluative bias.

The TPJ and STS are two regions associated with theory of mind and mentalizing. Drawing on previous work that examined mentalizing toward perceived Black and perceived White men who were targets of violence from law enforcement (e.g., Dang et al., 2022), we offer two main hypotheses. First, since depictions of violence tend to elicit greater mentalizing compared to non-violent scenes due to perceived intent to harm or personal relevance of the victim (Bruneau et al., 2012;

Young et al., 2011), we first predicted that the STS and TPJ would preferentially activate toward depictions of violence relative to the neutral images. Second, if differences in brain activity are observed across the two violent conditions, we predicted that participants may show greater activation of both the TPJ and STS to depictions of perceived Black victims relative to perceived White victims, replicating a previous investigation that found greater activation in either region to depictions of perceived Black men being violently arrested compared to perceived White men being violently arrested (Dang et al., 2022).

Lastly, while exploratory, we predicted that the VLPFC and particularly the DACC and DLPFC would show greater activity to depictions of violence relative to the neutral scenes. We predicted this pattern of results as these regions would likely reflect greater regulation of aversive responses to the less pleasant depictions of cross-race violence. If differences in activity did arise between the violent conditions of stimuli, we tentatively predicted to see greater activity in these regions to depictions of perceived Black victims relative to perceived White victims, as in order to more effectively mentalize toward perceived Black victims, greater response monitoring may also be necessary to override pro-ingroup evaluative biases.

Methods

Participants

The participants in Study 3 were identical to those in Study 2. All potential participants were screened prior to their experimental sessions to determine eligibility.

Eligible participants self-identified as White and non-Hispanic, were right-handed without a history of severe head injury or fainting, and otherwise met the University of Delaware's Center for Biomedical and Brain Imaging's (CBBI) safety criteria based on responses to a pre-study eligibility survey. Fifty-two participants were recruited from the University of Delaware and the surrounding community. One participant was excluded from the final analysis after indicating they did not self-identify as White and non-Hispanic on the post-scan demographics measure despite self-identifying on the pre-study eligibility survey. One additional participant was removed from the final analysis for excessive head movement, leaving a final sample of 49 participants (39 Women, 9 Men, 1 Genderfluid; M_{age} = 19.143, SD_{age} = 2.336). All participants provided informed consent in accordance with the guidelines set by the Declaration of Helsinki, and the study protocol was approved by the University of Delaware's Institutional Review Board.

Experimental Procedure

Study 3 used a procedure identical to that of Study 2.

Impression Formation Task

Similar to Study 2, participants then completed two runs of an impression formation task, where they viewed and passively formed an impression of the images of violent or neutral images. In this task, participants were instructed to form an impression of each image that they saw. Images were presented two times within the study, once per functional run. Each trial of the impression formation task consisted of

a centrally presented fixation cross displayed for 500ms, followed by image presentation for 2000ms. Following the presentation of the scene, participants would view a jittered fixation for intervals of 0, 2000, 4500, or 7000ms before the next trial would begin. Prior to the scan, participants were instructed to view and passively form impressions of the images of couples as they appeared on screen. They were tasked with pressing buttons corresponding to their thumbs on two button boxes they were given during the scan to ensure that they were actively attending to the images. Additionally, because their button presses did not differ for each trial, they were told that the impressions they formed were private. Each run consisted of a total of 60 trials, and runs lasted approximately 5 minutes.

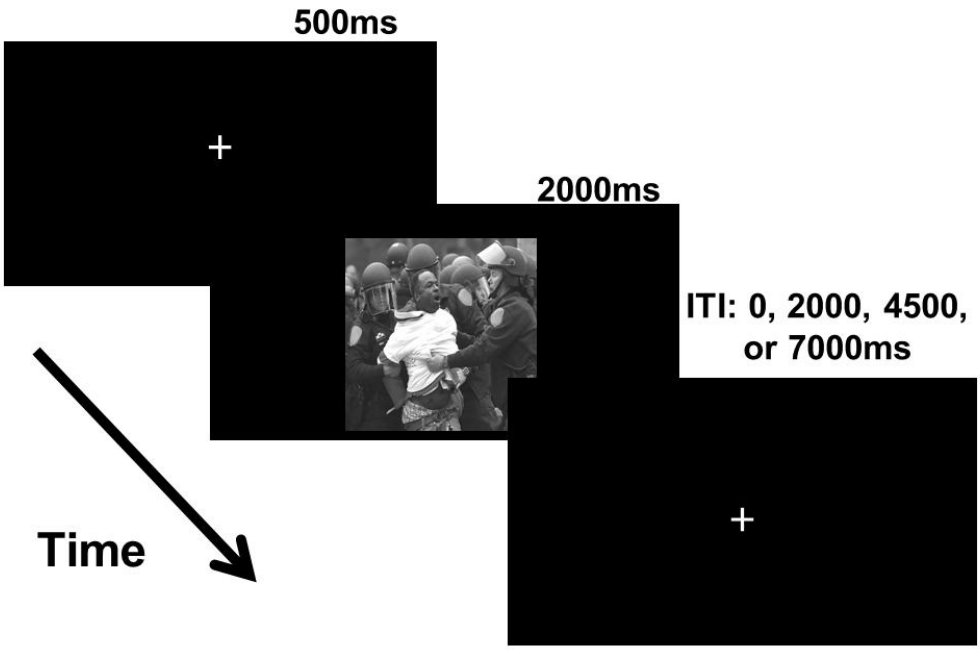

Figure 20. Schematic depicting trial structure for the impression formation task. In each trial, participants would view a centrally presented fixation cross for 500ms,

then would view and passively form an evaluation of the image during its 2000ms presentation window. A jittered intertrial fixation cross (ITI) would then be presented for 0, 2000, 4500, or 7000 ms.

Stimuli

The images used in Study 3 depicted violent scenes that were piloted and normed during Study 1B of this dissertation.

(A)

(B)

(C)

Figure 21. Example images depicting perceived Black victim(s) (Panel A), Neutral Scenes (Panel B), and perceived White (Panel C) victims.

fMRI Data Acquisition

MRI data were acquired on a 3T Siemens Magneton Prisma MRI scanner with a 64-channel head coil (Siemens Medical, Erlangen, Germany) at the University of Delaware Center for Biomedical and Brain Imaging. For each subject, functional

images were collected in 2 functional runs of 115 TRs each, interleaved in 38 axial slices (3.6mm thick) with a T2*-weighted echo-planar imaging (EPI) sequence: MB acceleration factor = 4, repetition time (TR) = 2500 ms, echo time (TE) = 27 ms, flip angle (FA) = 70 degrees, voxel size = 2.6 mm * 2.6 mm * 3.6 mm, field of view (FOV) = 230 mm * 230 mm. High-resolution structural images were acquired in the sagittal plane using a T1-weighted 3D Turbo Field Echo (TFE/MP-Rage) anatomical scan (TR = 2080ms, echo time = 4.64ms, FOV = 210 × 210 mm, slice thickness = 0.70 mm, voxel size = 0.70 × 0.70 × 0.70 mm).

Slices were aligned to the anterior commissure—posterior commissure axis of each participant. Each run lasted approximately 5 minutes. Timing sequences for each run were created using optseq2 (Greve, 2002). Each sequence was designed to be distinct and ensured that a stimuli condition was not presented more than three consecutive times. Sequence order was counterbalanced across participants. Functional imaging data were pre-processed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) and SPM12W (https://github.com/wagner-lab/spm12w) to remove noise and to correct for motion artifacts. The images were then transformed into a standard anatomical space (3mm isotropic voxels) based on the ICBM 152 template (MNI, Montreal Neurological Institute). Normalized data were then spatially smoothed using an 8mm full width at half maximum (FWHM) Gaussian Kernal to increase the signal-to-noise ratio and to reduce the impact of anatomical variability not corrected for by stereotaxic normalization. GLMs were

constructed for each participant to examine condition-specific brain activity as a function of perceived victim race (i.e., Black victim, White victim, Neutral).

Data Analysis

ROI fMRI data were analyzed using the LME4 package (Bates et al., 2015) in the R programming language (R Core Team, 2017). Analyses consisted of a within-subjects design and included an intercept for each subject. The within-subjects independent factor was the victim's perceived race. Three models were run to compare across all three levels of the within-subjects variable of the victim's perceived race. The first model compared perceived White victims to perceived Black victims and Neutral scenes (contrast code: -0.33 = perceived Black victim, -0.33 = Neutral scene, 0.66 = perceived White victim) and additionally compared perceived Black victims directly to Neutral scenes (0.5 = Neutral scenes, -0.5 = perceived Black victims). The second model compared Neutral scenes to perceived Black and perceived White victims (-0.33 = perceived Black victims, -0.33 = perceived White victims, 0.66 = Neutral scenes and additionally compared perceived White victims to perceived Black victims (0.5 perceived White victims, -0.5 = perceived Black victims). The third model perceived Black victims to perceived White victims and Neutral scenes (-0.33 = Neutral scenes, -0.33 = perceived White victims, 0.66 = perceived Black victims), additionally compared perceived White victims to Neutral scenes (0.5 = perceived White victims, -0.5 = Neutral scenes).

ROI Analysis

In Study 3, we examined ROIs implicated in the salience detection and negative evaluative responses, such as the amygdala and AI, those involved in mentalizing, such as the DMPFC, STS, and TPJ, as well as exploratory analysis on regions implicated in the regulation of intergroup prejudice, such as the DACC, DLPFC, and VLPFC. Bilateral amygdalae ROIs were modeled based on a previous investigation that examined race and status-based social evaluations left amygdala ($MNI_{x,y,z}$ = -24, -3, -21, 4 mm sphere), right amygdala ($MNI_{x,y,z}$ = 24, -3, -21, 4 mm sphere; (Mattan, Kubota, Dang, et al., 2018). Bilateral AI and DACC ROIs were modeled based on existing meta-analyses that examined the AI and DACC's role in salience detection and emotional regulation left AI ($MNI_{x,y,z}$ = -42, 16, 0, 8 mm sphere), right AI ($MNI_{x,y,z}$ = 36, 18, 0, 8 mm sphere; Adolfi et al., 2017), left DACC ($MNI_{x,y,z}$ = -4, 20, 38, 8 mm sphere), right DACC ($MNI_{x,y,z}$ = 6, 26, 38, 8 mm sphere; Zhang & Peng, 2023). Bilateral DLPFC and VLPFC coordinates were derived from a meta-analysis examining affective regulation left DLPFC ($MNI_{x,y,z}$ = -42, 9, 48, 8 mm sphere), right DLPFC ($MNI_{x,y,z}$ = 39, 21, 45, 8 mm sphere), left VLPFC ($MNI_{x,y,z}$ = -48, 30, -8, 8 mm sphere), right VLPFC ($MNI_{x,y,z}$ = 48, 30, -8, 8 mm sphere; (Denny et al., 2023). The DMPFC, bilateral STS, and bilateral TPJ coordinates were derived from previous investigations and reviews on intergroup mentalizing DMPFC ($MNI_{x,y,z}$ = -3, 55, 23, 8 mm sphere), left STS ($MNI_{x,y,z}$ = -59, -15, -16, 8 mm sphere), right STS ($MNI_{x,y,z}$ = 57, -10, -20, 8 mm sphere), left TPJ ($MNI_{x,y,z}$ = -56, -55, 26, 8 mm sphere),

and right TPJ ($MNI_{x,y,z}$ = 54, -51, 17, 8 mm sphere; Dang et al., 2022; Handley et al., 2022, 2023; Spreng et al., 2009).

We examined the beta weights (generated from the GLMs) of all ROIs as a function of the victim condition. For all ROI analyses, we were only able to model a random intercept for each participant because there were too few trials per condition. Similar to Study 2, to account for multiple comparison concerns across confirmatory ROIs (n = 4), a Bonferroni correction was applied to each confirmatory ROI with an adjusted $\alpha = 0.0125$. Again, while all significant results for both confirmatory and exploratory regions are reported, effects from confirmatory ROIs that do not survive the Bonferroni correction should be interpreted cautiously.

Whole-Brain Analysis

Complementary to the ROI analyses, we conducted an exploratory whole-brain analysis on our main contrasts of interest: White victims > Black victims, White victims > Neutral scenes, and Black victims > Neutral scenes. We used an uncorrected voxel level of $p < 0.001$ and a cluster extent threshold of 15 voxels. All whole-brain analyses should be treated as exploratory and interpreted in the context of the ROI results.

Results

Button Press Response Time

To ensure that participants were not actively attending to one stimulus condition more rapidly than another, we analyzed button press response times across the violence conditions using the LME4 package in R (Bates et al., 2015) across three models. The first model compared perceived White victims to perceived Black victims and Neutral scenes (contrast code: -0.33 = perceived Black victim, -0.33 = Neutral scene, 0.66 = perceived White victim) and additionally compared perceived Black victims directly to Neutral scenes (0.5 = Neutral scenes, -0.5 = perceived Black victims). The second model compared Neutral scenes to perceived Black and perceived White victims (-0.33 = perceived Black victims, -0.33 = perceived White victims, 0.66 = Neutral scenes and additionally compared perceived White victims to perceived Black victims (0.5 perceived White victims, -0.5 = perceived Black victims). The third model perceived Black victims to perceived White victims and Neutral scenes (-0.33 = Neutral scenes, -0.33 = perceived White victims, 0.66 = perceived Black victims), additionally compared perceived White victims to Neutral scenes (0.5 = perceived White victims, -0.5 = Neutral scenes).

All response times were log-transformed prior to analysis, and all trials with no button responses were removed prior to analysis. Results revealed no significant differences between violent conditions, $p > 0.229$ (M_{Black} = 1174.422ms, SD_{Black} =

423.167ms, $M_{Neutral}$ = 1189.619ms, $SD_{Neutral}$ = 413.355ms, M_{White} = 1170.193ms, SD_{White} = 402.201ms).

Confirmatory ROI

Left amygdala

Results found significantly greater BOLD activity in the left amygdala to neutral stimuli relative to perceived White victims, b = -0.053, SE = 0.025, $CI_{95\%}$ = [-0.102, -0.003], $t(96)$ = -2.07, p = 0.041, though this did not survive Bonferroni correction (see Figure 22). All other effects were not significant, $p > 0.074$.

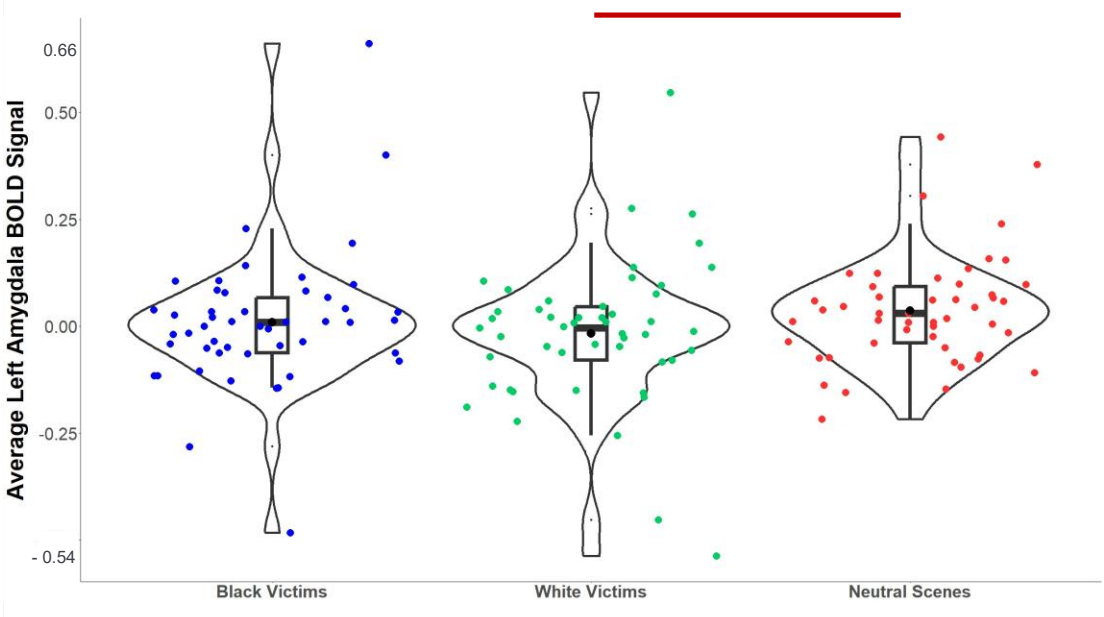

Figure 22. BOLD activity in the Left amygdala as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar

(black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. However, this effect did not survive Bonferroni correction.

Right amygdala

Results found greater activity to neutral relative to either condition of violent stimuli, $b = 0.042$, $SE = 0.0201$, $CI_{95\%} = [0.001, 0.083]$, $t(96) = 2.01$, $p = 0.048$, as well as greater activity toward neutral stimuli relative to depictions of perceived White victims, $b = -0.051$, $SE = 0.024$, $CI_{95\%} = [-0.097, -0.004]$, $t(96) = -2.14$, $p = 0.035$. However, neither effect survived Bonferroni correction (see Figure 23). All other effects were not significant, $p > 0.094$.

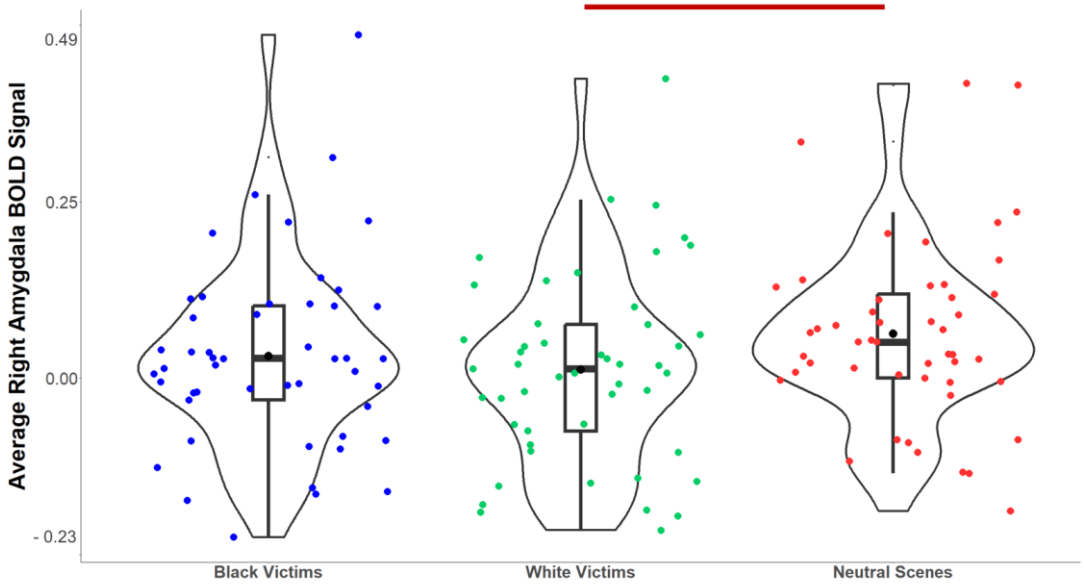

Figure 23. BOLD activity in the Right amygdala as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. However, this effect did not survive Bonferroni correction.

Left AI

Results in the left AI first found greater activity to perceived non-White victims, relative to perceived White ones, b = -0.267, SE = 0.047, $CI_{95\%}$ = [-0.359, -0.176], $t(96)$ = -5.75, $p < 0.001$. This effect was qualified by greater activity toward perceived Black relative to perceived White victims, b = -0.217, SE = 0.053, $CI_{95\%}$ = [-0.321, -0.113], $t(96)$ = -4.08, $p < 0.001$, and greater activity to neutral images compared to depictions of perceived White victims[1], b = -0.312, SE = 0.053, $CI_{95\%}$ = [-0.417, -0.208], $t(96)$ = -5.87, $p < 0.001$ (see Figure 24). Lastly, there was significantly greater activity to the neutral relative to the violent images, b = 0.206, SE

[1] To address concerns about outliers driving the significantly greater left AI activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, b = -0.273, SE = 0.040, $CI_{95\%}$ = [-0.364, -0.182], $t(94)$ = -5.906, $p < 0.001$.

= 0.046, $CI_{95\%}$ = [0.115, 0.297], $t(96) = 4.43$, $p < 0.001$. All other effects were not significant, $p > 0.0762$.

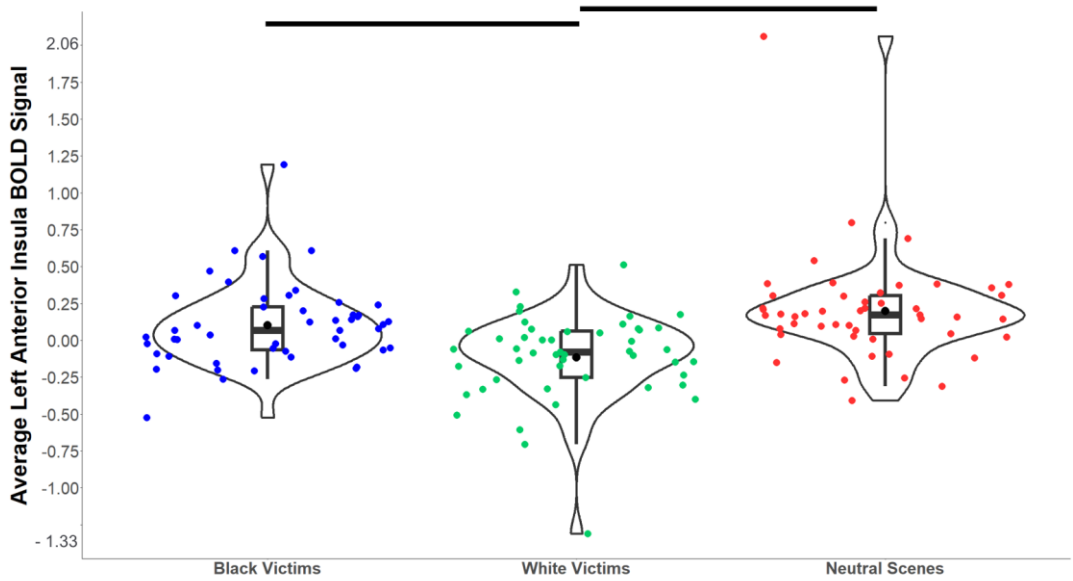

Figure 24. BOLD activity in the Left AI as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. Results also revealed significantly greater BOLD signal to scenes depicting perceived Black victims relative to those depicting perceived White victims.

Right AI

Results in the right AI found significantly lower BOLD activity to depictions of perceived White victims relative to perceived Black victims or neutral images, b = -0.097, SE = 0.030, $CI_{95\%}$ = [-0.156, -0.038], $t(96)$ = -3.24, p = 0.002. This effect was further qualified by significant differences in activity to depictions of perceived Black victims relative to perceived White victims, b = -0.091, SE = 0.0344, $CI_{95\%}$ = [-0.158, -0.023], $t(96)$ = -2.64, p = 0.01 and neutral images relative to depictions of perceived White victims, b = -0.102, SE = 0.034, $CI_{95\%}$ = [-0.17, -0.035], $t(96)$ = -2.97, p = 0.004 (see Figure 25). All other effects were not significant, p > 0.06.

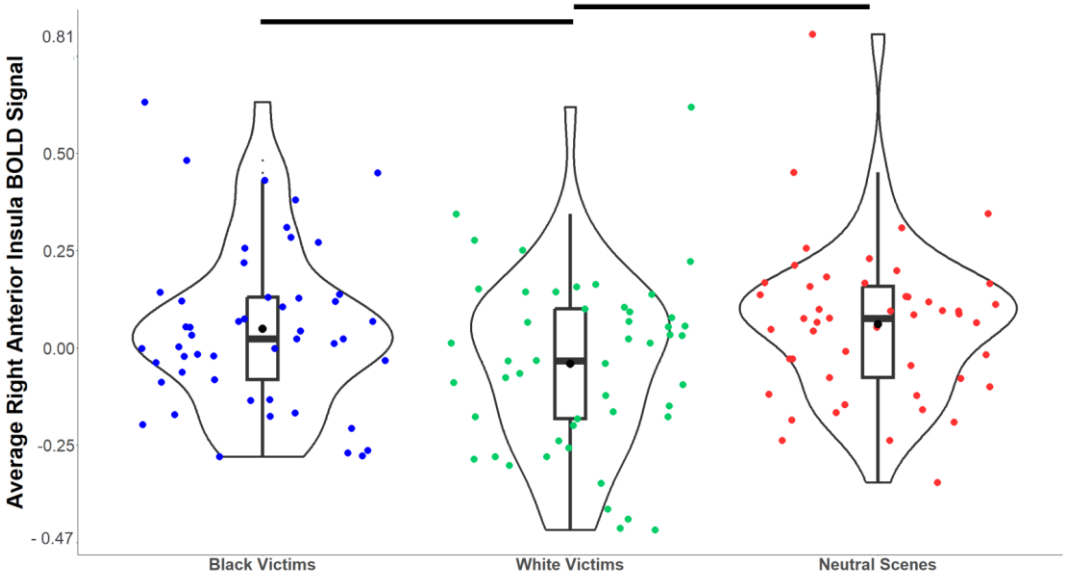

Figure 25. BOLD activity in the Right AI as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e.,

responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. Results also revealed significantly greater BOLD signal to scenes depicting perceived Black victims relative to those depicting perceived White victims.

Left TPJ

Results in the TPJ found significantly diminished activity toward perceived White victims relative to depictions of perceived Black victims or neutral scenes, b = -0.134, SE = 0.046, $CI_{95\%}$ = [-0.224, -0.044], $t(96)$ = -2.91, p = 0.004. This effect was qualified by significantly greater activity toward depictions of perceived Black victims relative to perceived White ones, b = -0.143, SE = 0.053, $CI_{95\%}$ = [-0.246, -0.040], $t(96)$ = -2.71, p = 0.008 (see Figure 26). There was additionally greater activity to neutral images relative to those depicting perceived White victims[2]. However, this effect did not survive Bonferroni correction, b = -0.123, SE = 0.053, $CI_{95\%}$ = [-0.226, -0.020], $t(96)$ = -2.34, p = 0.022. All other effects were not significant, $p > 0.078$.

[2] To address concerns about outliers driving the significantly greater left TPJ activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, b = -0.120, SE = 0.051, $CI_{95\%}$ = [-0.220, -0.199], $t(94)$ = -2.349, p = 0.021.

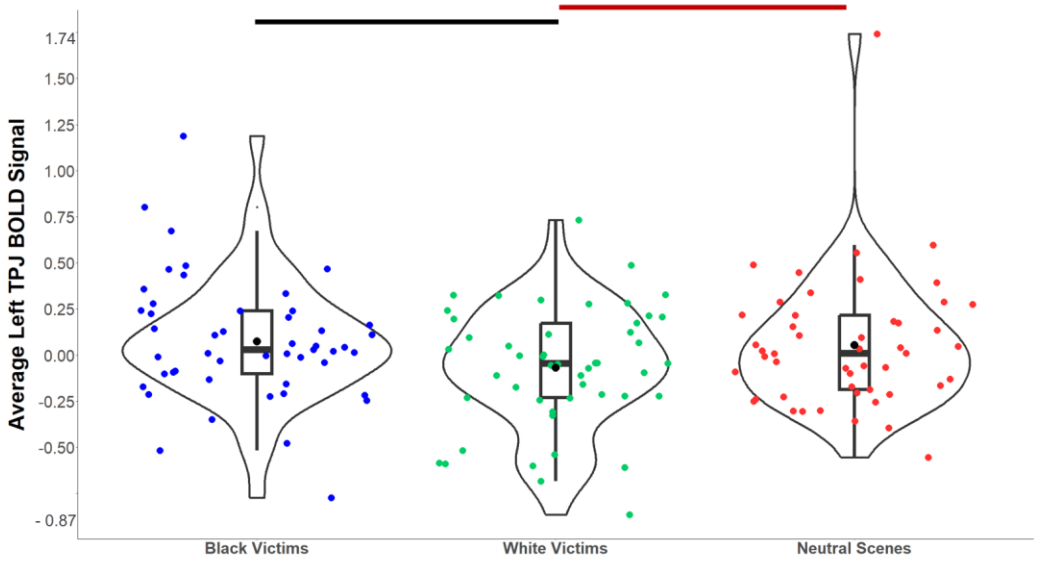

Figure 26. BOLD activity in the Left TPJ as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to scenes depicting perceived Black victims relative to scenes depicting perceived White victims. Moreover, results revealed greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. However, this effect did not survive Bonferroni correction.

Right TPJ

Results in the right TPJ found greater activity toward perceived Black victims relative to perceived White victims or neutral scenes, $b = 0.108$, $SE = 0.045$, $CI_{95\%} = [0.019, 0.197]$, $t(96) = 2.39$, $p = 0.019$. This finding was driven by greater activity to perceived Black relative to neutral images, $b = 0.113$, $SE = 0.052$, $CI_{95\%} = [0.0119,$

0.215], $t(96) = 2.19$, $p = 0.031$, however neither effect survived Bonferroni correction (see Figure 27). All other effects were not significant, $p > 0.055$.

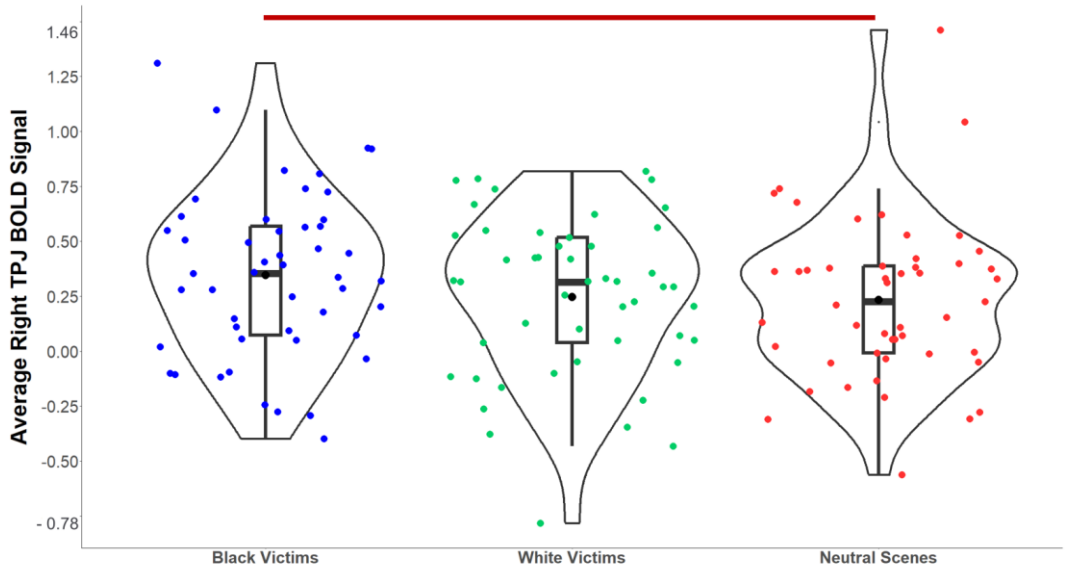

Figure 27. BOLD activity in the Right TPJ as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to scenes that depicted perceived Black victims relative to those that depicted neutral scenes. However, this effect did not survive Bonferroni correction.

Left STS

Results found greater activity to neutral images relative to either perceived Black or perceived White victims, $b = 0.101$, $SE = 0.031$, $CI_{95\%} = [0.040, 0.161]$, $t(96)$

= 3.25, $p = 0.002$. This effect was driven by greater activity to neutral relative to perceived White victims[3], $b = -0.129$, $SE = 0.035$, $CI_{95\%} = [-0.198, -0.059]$, $t(96) = -3.63$, $p<0.001$. Additionally, this was driven by greater activity to neutral scenes relative to perceived Black victims[4], $b = -0.071$, $SE = 0.035$, $CI_{95\%} = [-0.140, -0.001]$, $t(96) = -1.99$, $p = 0.049$. However, this effect did not survive Bonferroni correction. Additionally, there was significantly diminished activity to perceived White victims relative to depictions of perceived Black victims or neutral scenes, $b = -0.094$, $SE = 0.031$, $CI_{95\%} = [-0.155, -0.034]$, $t(96) = -3.04$, $p = 0.003$ (see Figure 28). All other effects were not significant, $p > 0.104$.

[3] To address concerns about outliers driving the significantly greater left STS activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, $b = -0.130$, SE = 0.035, $CI_{95\%} = [-0.200, -0.061]$, $t(94) = -3.666$, $p < 0.001$.

[4] Removal of the outlier responses did however, cause the differences in left STS activity toward the neutral scenes to no longer significantly differ from the scenes depicting Black victims, $p = 0.078$.

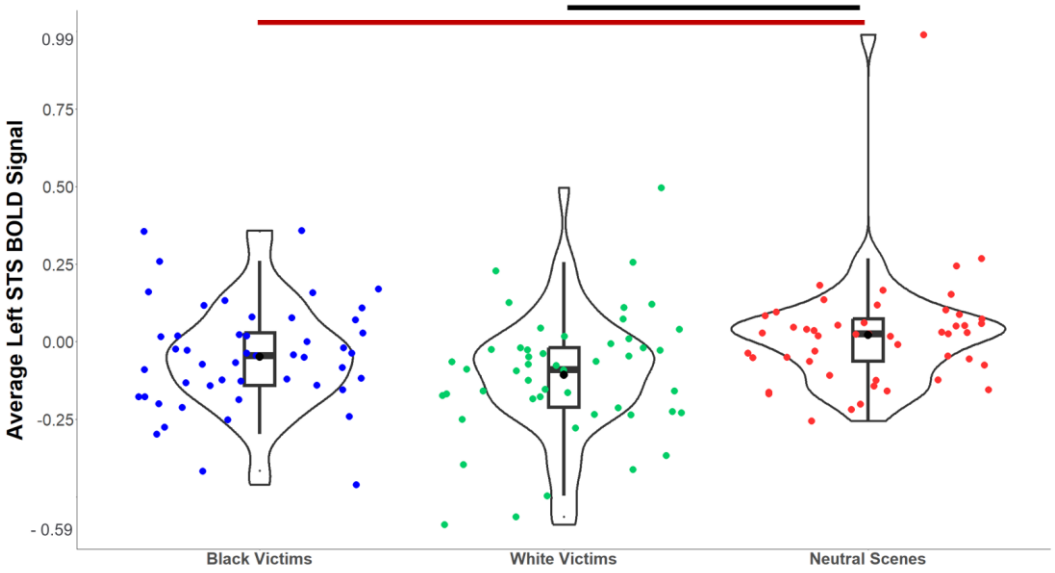

Figure 28. BOLD activity in the Left STS as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. Additionally, results revealed significantly greater BOLD signal to neutral scenes relative to those that depicted perceived Black victims. However, this effect did not survive Bonferroni correction.

Right STS

Results in the right STS found greater activity toward neutral images relative to either violent condition, b = 0.084, SE = 0.025, $CI_{95\%}$ = [0.035, 0.133], $t(96)$ = 3.36, p = 0.001. This effect was driven by greater activity to neutral images relative to those depicting a perceived White victim, b = -0.097, SE = 0.029, $CI_{95\%}$ = [-0.155, -0.043],

$t(96) = -3.44$, $p<0.001$. Additionally, there was greater activity toward neutral images compared to those that depicted a perceived Black victim[5], $b = -0.068$, $SE = 0.029$, $CI_{95\%} = [-0.124, -0.0121]$, $t(96) = -2.38$, $p = 0.019$. However, this effect did not survive Bonferroni correction (see Figure 29). Lastly, there was significantly decreased activity to images depicting perceived White victims relative to those depicting perceived Black victims or neutral scenes, $b = -0.065$, $SE = 0.025$, $CI_{95\%} = [-0.114, -0.016]$, $t(96) = -2.6$, $p = 0.011$. All other effects were not significant, $p > 0.291$.

[5] To address concerns about outliers driving the significantly greater right STS activity to neutral scenes relative to those depicting Black victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, $b = -0.057$, SE = 0.027, $CI_{95\%} = [-0.111, -0.004]$, $t(94) = -2.113$, $p = 0.037$.

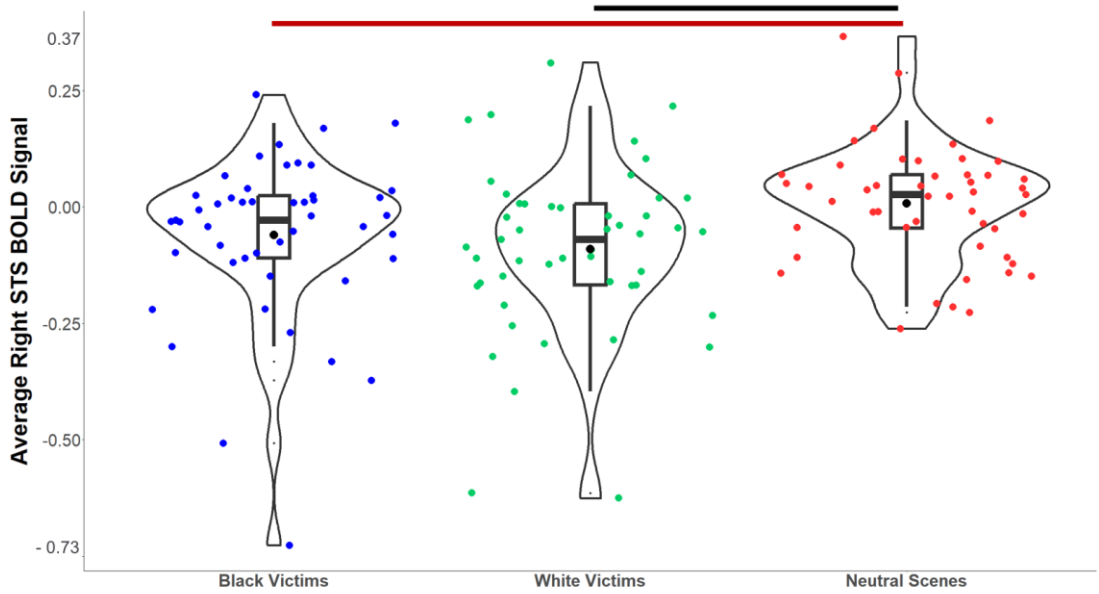

Figure 29. BOLD activity in the Right STS as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. perceived Black lines indicate an effect that survived Bonferroni correction, whereas red lines indicate the significant effect did not survive Bonferroni correction. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. Additionally, results revealed significantly greater BOLD signal to neutral scenes relative to those that depicted perceived Black victims. However, this effect did not survive Bonferroni correction.

Exploratory ROI

Left DACC

Results in the left DACC found significantly lower activity to depictions of perceived White victims relative to neutral scenes or perceived Black victims, b = -

0.10, *SE* = 0.034, $CI_{95\%}$ = [-0.167, -0.033], *t*(96) = -2.91, *p* = 0.004. This effect was further qualified by greater activity toward neutral scenes relative to perceived White victims, *b* = -0.121, *SE* = 0.039, $CI_{95\%}$ = [-0.198, -0.044], *t*(96) = -3.08, *p* = 0.003 (see Figure 30). Additionally, there was significantly greater activity toward neutral images compared to either condition of violent image *b* = 0.083, *SE* = 0.034, $CI_{95\%}$ = [0.0161, 0.15], *t*(96) = 2.43, *p* = 0.017. All other effects were not significant, *p* > 0.053.

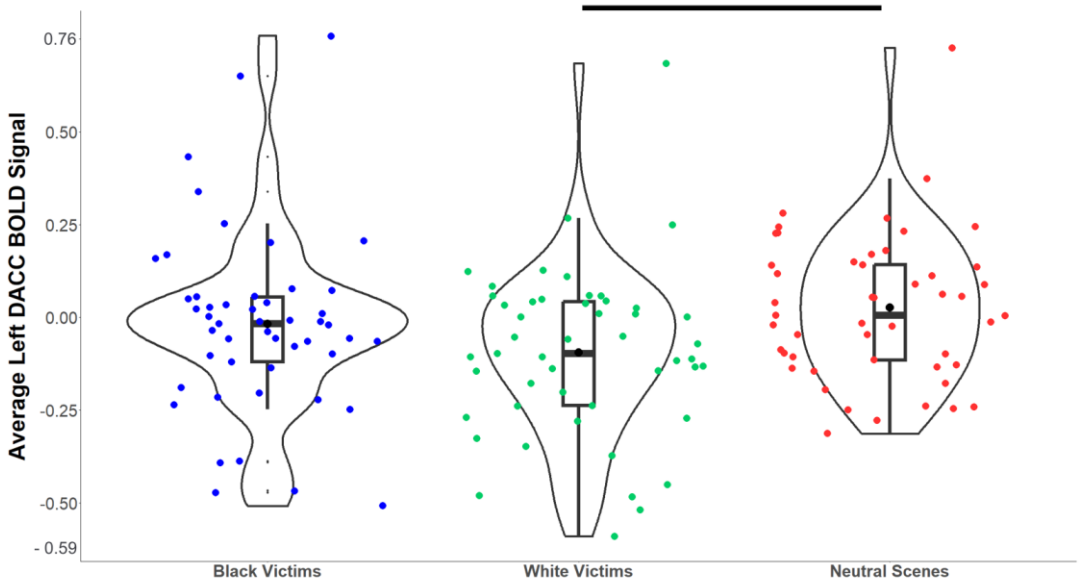

Figure 30. BOLD activity in the Left DACC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims.

Right DACC

Results in the right DACC found greater activity to neutral images relative to the violent images, $b = 0.094$, $SE = 0.039$, $CI_{95\%} = [0.018, 0.170]$, $t(96) = 2.42$, $p = 0.017$. Additionally, there was greater BOLD activity to neutral images relative to those depicting a perceived White victim, $b = -0.117$, $SE = 0.044$, $CI_{95\%} = [-0.203, -0.030]$, $t(96) = -2.63$, $p = 0.010$ (see Figure 31). There was also greater activity toward images that did not depict a perceived White victim, $b = -0.083$, $SE = 0.039$, $CI_{95\%} = [-0.159, -0.006]$, $t(96) = -2.13$, $p = 0.036$. All other effects were not significant, $p > 0.120$.

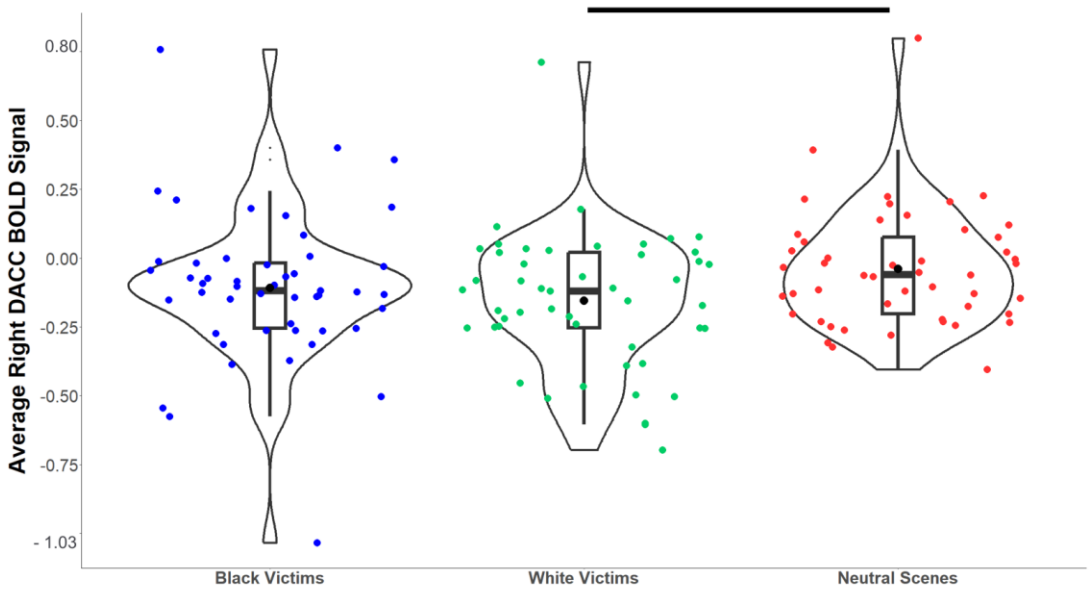

Figure 31. BOLD activity in the Right DACC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions

represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims.

Left DLPFC

Results found diminished BOLD activity to images depicting perceived White victims, relative to perceived Black victims or neutral scenes, b = -0.208, SE = 0.047, $CI_{95\%}$ = [-0.301, -0.115], $t(96)$ = -4.39, p< 0.001. This was qualified by greater activity toward violent images depicting perceived Black victims relative to perceived White victims, b = -0.172, SE = 0.054, $CI_{95\%}$ = [-0.278, -0.066], $t(96)$ = -3.17, p = 0.002, and neutral images relative to those depicting perceived White victims, b = -0.240, SE = 0.054, $CI_{95\%}$ = [-0.347, -0.134], $t(96)$ = -4.44, p< 0.001 (see Figure 32). Moreover, there was greater activity toward neutral images relative to either violent image condition, b = 0.156, SE = 0.047, $CI_{95\%}$ = [0.063, 0.249], $t(96)$ = 3.29, p = 0.001. All other effects were not significant, p > 0.210.

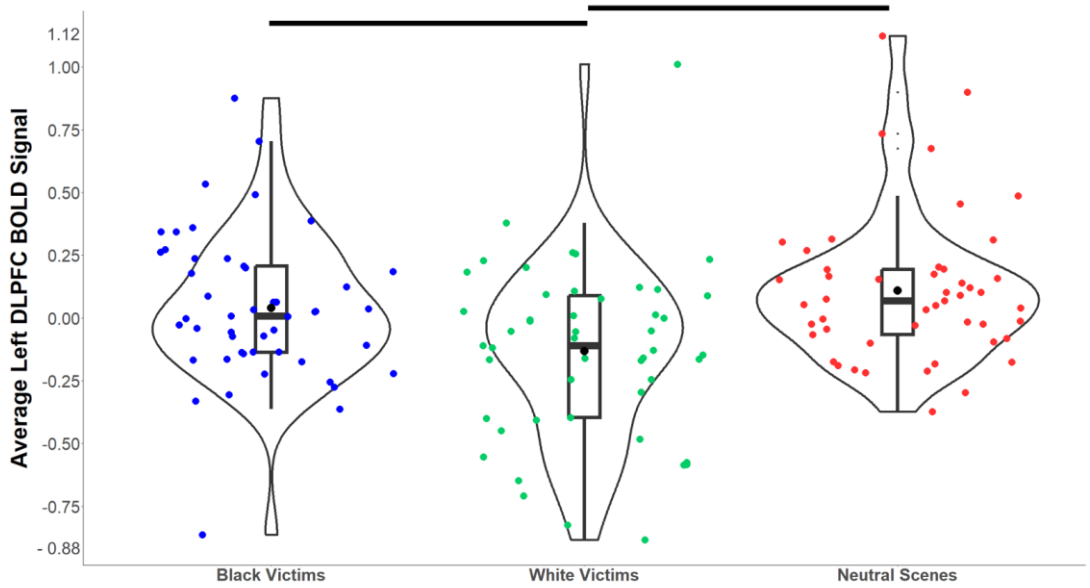

Figure 32. BOLD activity in the Left DLPFC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims. Additionally, results revealed significantly greater BOLD signal to scenes that depicted perceived Black victims relative to those that depicted perceived White victims.

Right DLPFC

Results in the right DLPFC revealed greater activity to Neutral images relative to either violent image condition, $b = 0.146$, $SE = 0.0511$, $CI_{95\%} = [0.046, 0.246]$, $t(96) = 2.85$, $p = 0.005$. Additionally, results found diminished activity to depictions of a perceived White victim, $b = -0.117$, $SE = 0.051$, $CI_{95\%} = [-0.217, -0.017]$, $t(96) = -2.29$, $p = 0.024$, which appeared to be driven by greater activity toward neutral relative

to perceived White victims[6], b = -0.173, SE = 0.058, $CI_{95\%}$ = [-0.288, -0.059], $t(96)$ = -2.97, p = 0.004 (see Figure 33). All other effects were not significant, $p > 0.052$.

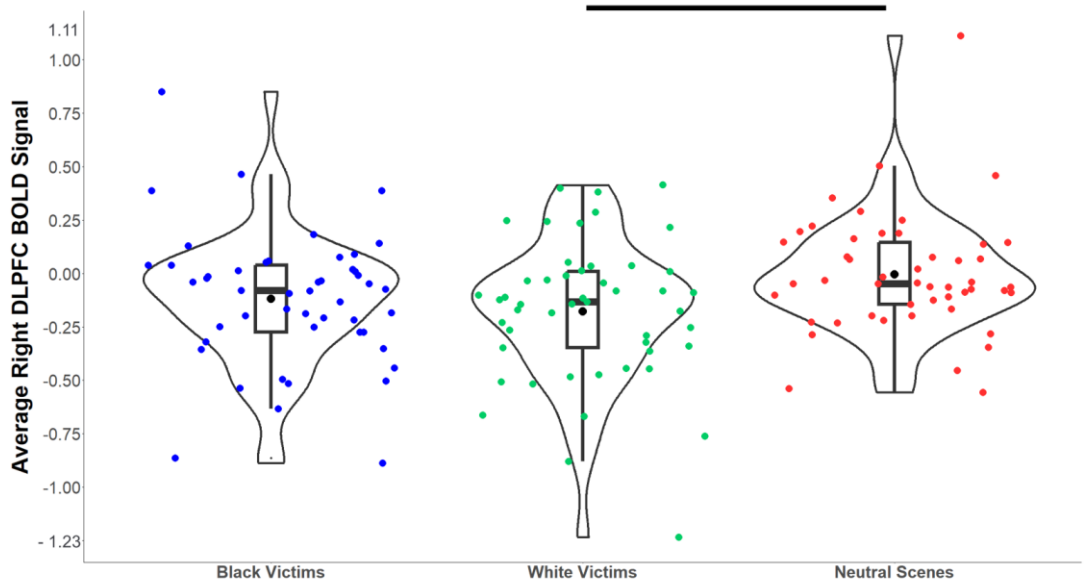

Figure 33. BOLD activity in the Right DLPFC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims.

[6] To address concerns about outliers driving the significantly greater right DLPFC activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, b = -0.128, SE = 0.049, $CI_{95\%}$ = [-0.224, -0.032], $t(94)$ = -2.602, p = 0.011.

Left VLPFC

Results revealed greater activity to the neutral relative to violent stimuli conditions, $b = 0.164$, $SE = 0.037$, $CI_{95\%} = [0.091, 0.236]$, $t(96) = 4.41$, $p<0.001$. This was qualified by greater activity to neutral images relative to those that depicted a perceived White victim, $b = -0.147$, $SE = 0.042$, $CI_{95\%} = [-0.230, -0.064]$, $t(96) = -3.47$, $p<0.001$, or a perceived Black victim[7], $b = -0.176$, $SE = 0.042$, $CI_{95\%} = [-0.260, -0.093]$, $t(96) = -4.16$, $p<0.001$ (see Figure 34). Moreover, there was significantly diminished activity to depictions of perceived Black victims relative to the other two conditions, $b = -0.104$, $SE = 0.037$, $CI_{95\%} = [-0.177, -0.031]$, $t(96) = -2.8$, $p = 0.006$. All other effects were not significant, $p > 0.111$.

[7] To address concerns about outliers driving the significantly greater left VLPFC activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting Black victims, $b = -0.156$, SE = 0.040, $CI_{95\%} = [-0.235, -0.077]$, $t(94) = -3.875$, $p < 0.001$.

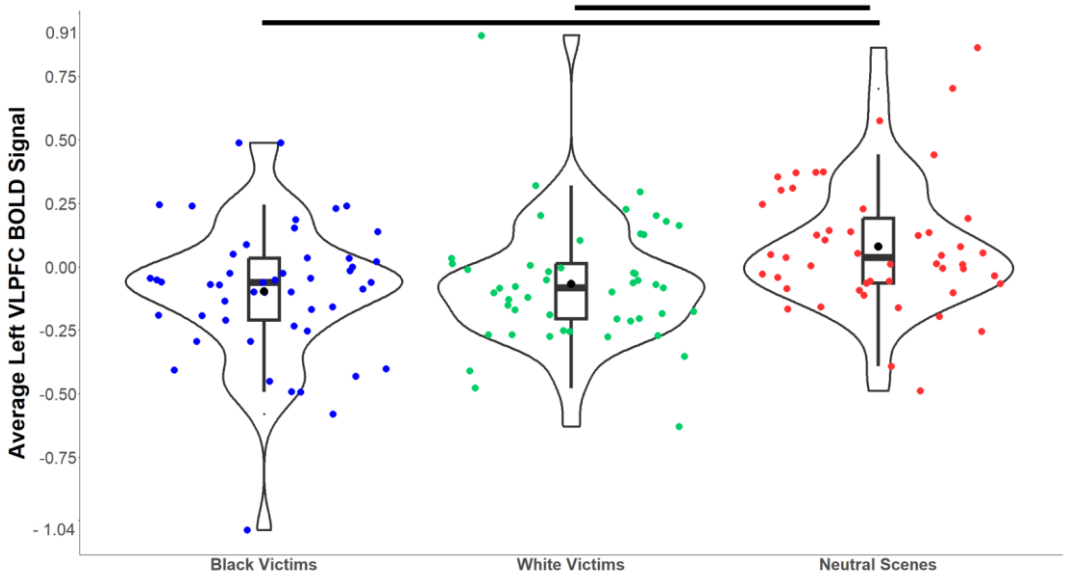

Figure 34. BOLD activity in the Left VLPFC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to both scenes that depicted perceived Black victims as well as those that depicted perceived White victims.

Right VLPFC

Results in the right VLPFC revealed significantly greater activity to neutral stimuli relative to either violent condition, $b = 0.12$, $SE = 0.037$, $CI_{95\%} = [0.048, 0.192]$, $t(96) = 3.25$, $p = 0.002$. This was further qualified by greater activity to neutral

images relative to depictions of perceived White victims[8], b = -0.122, SE = 0.042, $CI_{95\%}$ = [-0.204, -0.039], $t(96)$ = -2.89, p = 0.005, or perceived Black victims[9], b = -0.115, SE = 0.042, $CI_{95\%}$ = [-0.198, -0.033], $t(96)$ = -2.74, p = 0.007 (see Figure 35). All other effects were not significant, $p > 0.082$.

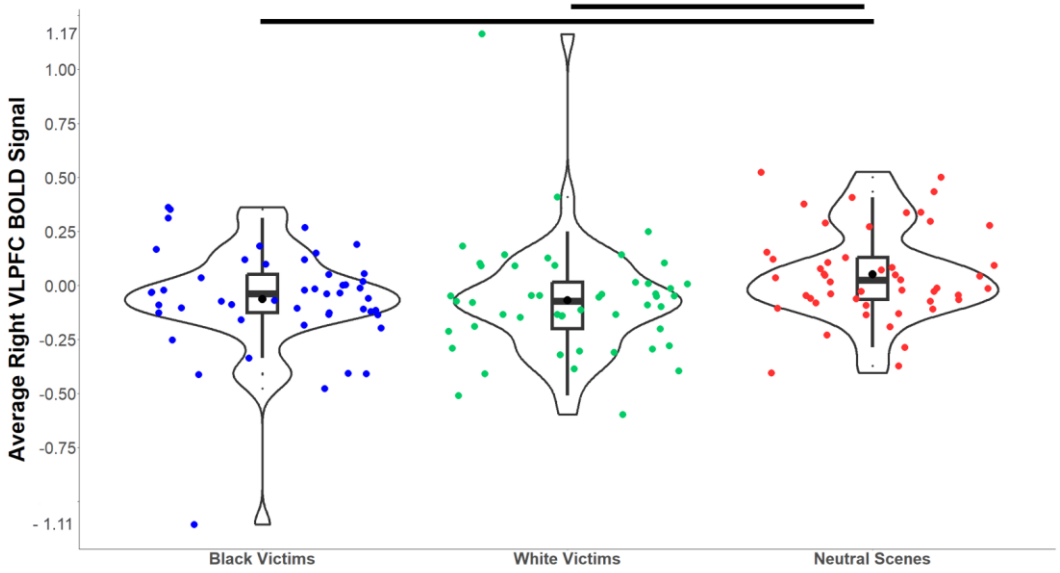

Figure 35. BOLD activity in the Right VLPFC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean

---

[8] To address concerns about outliers driving the significantly greater right VLPFC activity to neutral scenes relative to those depicting White victims, follow-up models were run with the outliers included. Results still yielded significantly greater BOLD activity to the neutral scenes relative to the scenes depicting White victims, b = -0.139, SE = 0.037, $CI_{95\%}$ = [-0.212, -0.067], $t(92)$ = -2.602, $p < 0.001$.

[9] Additionally results revealed there was significantly greater activity in right VLPFC activity to neutral scenes relative to those depicting Black victims, b = -0.091, SE = 0.037, $CI_{95\%}$ = [-0.164, -0.018], $t(92)$ = -2.453, p = 0.016.

and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to both scenes that depicted perceived Black victims as well as those that depicted perceived White victims.

DMPFC

Results in the DMPFC found greater activity toward neutral images relative to either other condition, $b = 0.179$, $SE = 0.089$, $CI_{95\%} = [0.003, 0.354]$, $t(96) = 2$, $p = 0.049$. This effect was qualified by greater activity to neutral images relative to those depicting perceived White victims, $b = -0.27$, $SE = 0.102$, $CI_{95\%} = [-0.471, -0.070]$, $t(96) = -2.65$, $p = 0.010$ (see Figure 36). Moreover, there was significantly greater activity to images that did not depict perceived White victims, $b = -0.231$, $SE = 0.089$, $CI_{95\%} = [-0.406, -0.056]$, $t(96) = -2.59$, $p = 0.011$. All other effects were not significant, $p > 0.070$.

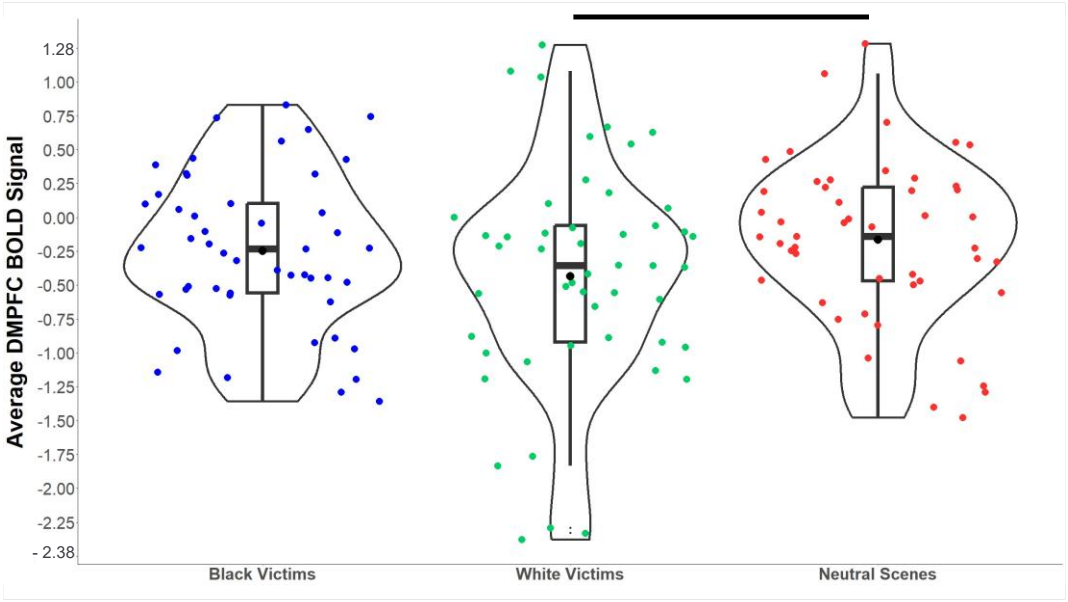

Figure 36. BOLD activity in the DMPFC as a function of victim race. Blue dots represent depictions of perceived Black victims, green dots represent depictions of perceived White victims, and red dots represent depictions of neutral scenes. The y-axis represents average BOLD activity, and the x-axis represents conditions. The mean and the 95% confidence interval are displayed as a point estimate and vertical bar (black dot and line, respectively). The boxes signify the interquartile responses (i.e., responses in the 25th and 75th percentile of these data). The lines between conditions represent significant differences as a function of experimental condition. Results revealed significantly greater BOLD signal to neutral scenes relative to scenes that depicted perceived White victims.

Whole-Brain

Table 8 Whole-brain bold signal as a function of Victim Race.

					MNI Coordinates		
Contrast	**Brain Region**	**BA**	***K***	***t-value***	***x***	***y***	***z***
Black victims > Neutral Scenes	L.Middle Temporal Gyrus		757	7.35	-42	-72	15
	R.Middle Temporal Gyrus		844	6.88	45	-63	6

	R.Middle Occipital Gyrus			6.16	27	-84	15
Black victims > White victims							
	L.Insula		575	5.12	-36	15	3
	L.Inferior Frontal Gyrus	44		4.75	-54	3	18
	L.superior Frontal gyrus		102	5.06	-18	48	39
	Brainstem		83	4.91	-9	-27	-12
	L.Supplemental Motor Area	6	268	4.65	-9	-12	69
	R.Superior Frontal Gyrus			4.22	15	6	69
	R.Insula		130	4.58	42	9	3
	R.Inferior Frontal Gyrus			4.10	48	18	0
	R.Postcentral Gyrus		37	4.47	66	-21	24
	L.Cuneus		77	4.31	-15	-87	27
	R.Postcentral Gyrus		267	4.18	18	-42	72
	R.Precentral Gyrus			4.03	63	-21	45
	L.Inferior Parietal Lobule	40	243	4.02	-57	-36	48
	L.Postcentral Gyrus	40		3.97	-45	-36	60
	L.Inferior Parietal Lobule		24	3.72	-45	-48	27
	L.Superior Frontal Gyrus			3.49	-6	51	36
	L.Superior Temporal Gyrus		15	3.58	-54	-21	0
Neutral Scenes > Black victims	L.Uncus		272	5.38	-36	-9	-39
	R.Parahippocampal Gyrus	30	342	5.72	18	-39	-9
	R.Hippocampus			4.33	27	-39	0
	R.Cingulate Gyrus	24	125	4.48	6	-9	27

	L.Superior Frontal Gyrus	6	484	5.52	-9	-18	78
	R.Superior Frontal Gyrus			4.58	15	-9	72
	R.Superior Frontal Gyrus	6		4.46	18	-21	72
	R.Fusiform Gyrus		330	5.30	51	-15	-30
	R.Temporal Inferior Gyrus			4.95	45	-21	-24
	R.Parahippocampal Gyrus			4.93	33	-12	-30
	L.Inferior Temporal Gyrus	20	165	4.70	-54	-27	-30
	L.Fusiform Gyrus	20		4.27	-45	-24	-33
	L.Middle Temporal Gyrus			4.17	-51	-18	-21
	VMPFC	25	108	4.66	6	15	-9
	R.Medial Prefrontal Cortex	47		4.01	15	24	-18
	R.Precentral Gyrus		85	4.45	57	12	6
	R.Inferior Frontal Gyrus			4.27	60	12	21
	R.Superior Temporal Gyrus		36	4.26	51	-15	6
	L.Caudate		21	4.26	-9	15	9
	L.Inferior Parietal Lobule		57	4.26	-42	-45	51
	Cerebellum		19	4.07	0	-57	-48
	R.Middle Frontal Gyrus		21	4.04	-30	3	63
	R.Postcentral Gyrus		41	3.95	45	-36	51
	R.Inferior Partietal Lobule	40		3.57	42	-54	51
Neutral Scenes > White victims	L. Inferior Frontal Gyrus		158950	6.26	-45	30	12
	L.Caudate			5.91	-12	15	9
	R. Caudate			5.49	12	18	6
	R.Insula			5.45	45	-6	-3
	L.Insula			4.90	-42	9	0
	L.Hippocampus			4.78	-24	-39	0

	R.Hippocampus		4.34	24	-42	-3
	R.Inferior Frontal Gyrus		4.25	54	24	24
	R.Middle Temporal Gyrus	48	4.47	45	0	-39
	R.Pyramis	22	3.84	24	-63	-42
	Cuneus	35	3.68	0	-75	30
	R.Inferior Parietal Lobule	15	3.68	42	-54	48
	R.Culmen		3.66	33	-45	-33
White victims > Black victims	N/A					
White victims > Neutral Scenes	L.Middle Occipital Gyrus	360	6.42	-39	-75	12
	R.Middle Temporal Gyrus	535	6.34	42	-66	6
	R.Middle Occipital Gyrus		4.68	36	-84	18

Discussion

Study 3 examined how a sample of White perceivers evaluated depictions of violence toward either perceived Black or perceived White victims, as well as depictions of perceived White people during mundane social interactions. Results found there was greater AI activity to depictions of neutral scenes relative to either condition of violent stimuli. However, between either violent condition, results showed greater activity in the AI to depictions of perceived Black relative to perceived White victims. Additionally, results found greater activity in the STS and DMPFC to depictions of neutral social scenes relative to depictions of perceived White victims. Moreover, between violent conditions, there was greater activity in the TPJ to depictions of perceived Black relative to perceived White victims. Lastly, participants

showed greater activity in regions supporting regulation toward the depictions of neutral social scenes. We discuss these findings in further detail in the following paragraphs.

First, Study 3 found greater activity in the AI to depictions of perceived Black victims relative to the White victims. This possibly suggests that the AI may have been tracking participants' evaluations of the scenes, as between the violent conditions. Because this pattern of greater activity to perceived Black relative to White victims was only present in the AI and not the amygdala, this may tentatively mean that participants did find depictions of violence toward perceived Black people as more aversive than depictions toward perceived White people. Further, the AI activity being associated with greater aversion may be a plausible interpretation given its role in processing moral transgressions (e.g., Ying et al., 2018). Additionally, we found greater activity in the AI to depictions of neutral scenes relative to either condition of violent scene. This result was unexpected, given previous work establishing that the AI may be receptive to processing aversive stimuli (Borg et al., 2008; Duerden et al., 2013; Simmons et al., 2004; Straube & Miltner, 2011). One possible explanation for these findings is that depictions of violence are perhaps less salient than initially thought. However, it may likely be attributed to the content of the images, specifically the relative novelty of the individuals' behavior in neutral relative to the violent scenes. While all image conditions were equated on multiple dimensions, the violent conditions were generally very similar in the behavior and emotional expression of the individuals in each scene. In contrast, the neutral scenes had more variation in the

behavior of each person in the scene (e.g., men playing chess, exercising, playing cards). As such, it is possible that the neutral scenes may have been viewed as more novel to participants as there may have been more variety of information present for them to process in the neutral scenes relative to the violent scenes. Therefore, the greater activity in the AI may have been in support of more effortful salience processing of the neutral images (Menon & Uddin, 2010; Uddin, 2015).

Results from Study 3 also suggest that participants did show greater effort to mentalize with outgroup members, as they showed greater activity in the TPJ toward the perceived Black victims relative to the perceived White victims depicted in the violent stimuli. This finding does replicate existing work on intergroup mentalizing (Dang et al., 2022). Specifically, it replicates results that suggest greater mentalizing toward depictions of the use of force by police officers to perceived Black individuals relative to perceived White individuals. In this previous investigation, greater mentalizing toward perceived Black individuals during arrests was particularly enhanced when the level of force was not perceived to be fair relative to the arrestee's behavior. These scenes depict what may be viewed as unwarranted excessive force, particularly considering that many of the violent images depict law enforcement or members of the military as aggressors. Therefore, these findings may also be similar to Dang et al.'s findings, where participants mentalized with perceived Black victims more than perceived White victims considering the pervasiveness of police-sanctioned violence towards perceived Black people in the United States (e.g., Campbell & Valera, 2020; Canevez et al., 2022).

Results also revealed preferential activity in the STS and DMPFC to depictions of neutral scenes relative to those depicting perceived White victims. While the pattern of these results in either the STS or DMPFC should be interpreted with caution, it may again be attributed to the more ambiguous nature of the neutral scenes relative to the violent ones. The greater STS activity toward neutral scenes may be explained by the requirement of more effortful employment of the STS in efforts of participants to understand what was occurring in the neutral scenes as the greater variation in behaviors depicted in the neutral scenes may have made them more difficult to process (Isik et al., 2017). Further, since the STS is implicated in gleaning mental states from faces (Cloutier, Turk, et al., 2008; Handley et al., 2022, 2023; Hoffman & Haxby, 2000; Schurz et al., 2014), the greater ambiguity of the neutral scenes could have made gleaning the mental states of the individuals in each scene more difficult, and the greater STS activity to neutral scenes may have reflected greater effort needed to process them. However, if this is the case, it is not completely clear as to why there was only preferential STS activity to neutral scenes relative to the scenes depicting perceived White victims and not between scenes depicting perceived Black victims and those depicting neutral scenes.

Last, results found this sample of White perceivers showed greater DLPFC activity to perceived Black relative to perceived White victims. Because the DLPFC has been implicated in executive function related to intergroup bias regulation (Cassidy et al., 2016; Sreenivasan et al., 2014), these patterns of activity may mean that participants were more deliberately monitoring their own responses to the violent

images. The greater DLPFC activity to Black victims may have been in service of evaluating the scenes depicting perceived Black victims as aversive and the relatively more effortful mentalizing with the perceived Black victims. Results also found greater activity in the DACC, DLPFC, and VLPFC to depictions of neutral social scenes relative to either condition of violent stimuli. At first glance, these patterns of results are unexpected; however, they may be similarly attributed to the relative ambiguity of the neutral social scenes compared to either violent image condition. One possibility is that participants were engaging in executive control more in the interest of monitoring their responses toward the neutral scenes, and this process was more effortful due to the relative ambiguity of the behaviors of the individuals within the neutral scenes. Taken together, this suggests that while participants may have engaged more executive control when processing more complex social scenes when participants did show differences in regulation between violence conditions, as indexed by greater DLPFC activity, it was done so in the interest of monitoring responses to more effectively evaluate the scenes and mentalize with outgroup members.

Limitations

Similar to Study 2, Study 3 offered its own set of limitations due to the sample and study procedures. In the following paragraph we summarize these limitations and suggest future directions for subsequent investigations on how perceivers process aversive cross-race social interactions. Also, parallel to Study 2, Study 3's results may

be limited in their generalizability, given that the sample of 49 participants was primarily comprised of university undergraduate students. Similar to Study 2, Study 3 did not collect any index of motivation to regulate prejudice for every individual in the sample. Considering the pattern of results in the brain regions supporting regulation, measuring IMS/EMS may help better contextualize the results and may further explain differences in regulation between violent and non-violent conditions.

Relatedly, one of Study 3's biggest limitations may be the differences in the variability of behaviors between the violent and neutral scenes. Based on the results, it appears possible that many of the differences between the neutral and violent stimuli may indeed be attributed to the greater ambiguity of behaviors displayed in the neutral scenes. One way to perhaps better interpret these results would be to employ eye-tracking while participants process the violent and neutral scenes. This approach may help address concerns about the more ambiguous nature of the behaviors displayed in the neutral scenes, as this approach would allow for the tracking of salience detection through pupil size measurement (Mathôt, 2018) as well as attention to the scenes through saccadic deviation, or the extent to which participants fixate on the social scene or avert their gaze (Chakrabarti et al., 2017; Di Stasi et al., 2013) while also accounting for what parts of each scene participants are actively attending to. Still, despite the limitations, Study 3 provides more insights into how perceivers differentially process more complex social scenes, in this case, how they process more aversive, violent social scenes relative to scenes depicting mundane social interactions.

Conclusion

Study 3 explored the neural correlates of aversion/salience processing, intergroup mentalizing, and bias regulation during the processing of depictions of cross-race violence. In this study, we found that greater activity in the AI, a region supporting more aversive social evaluations to depictions of violence toward perceived Black people relative to depictions of violence toward perceived White people. Furthermore, Study 3's findings suggest that the neutral scenes may have been more ambiguous relative to either violent condition and, therefore, might have required more effort from participants to mentalize with the individuals in the neutral scenes. Depictions of violence toward perceived Black victims elicited more effortful mentalizing than depictions of violence toward perceived White victims, effectively partially replicating previous work on intergroup mentalizing. Lastly, exploratory ROI and whole-brain analyses revealed that regions supporting bias regulation may also be preferentially implicated in processing the more complex social scenes.

Chapter 6

GENERAL DISCUSSION

This thesis sought to examine how individuals differentially process depictions of social behaviors between members of their own ingroup (i.e., perceived White people) and outgroup members (i.e., perceived Black people) when the interactions depicted are either affiliative or aversive. In Study 2, while undergoing fMRI, a sample of White perceivers passively viewed and formed impressions of happy smiling couples, which were comprised of partners who differed in racial composition, as curated in Study 1A. Confirmatory results in Study 2 focused on brain regions supporting positive evaluation and salience/aversion, and exploratory results examined regions that are thought to support the regulation of intergroup bias. In Study 3, while undergoing fMRI, the same sample of White perceivers passively viewed and formed impressions of images that depicted mundane social interactions or depictions of a perceived White aggressor attacking a perceived Black or White victim that was curated in Study 1B. Confirmatory results in Study 3 were centered on brain regions that are thought to support salience/aversion and mentalizing toward others, and exploratory results centered on regions supporting the regulation of intergroup bias. In the following section, we review the current hypotheses and subsequent findings from Studies 2 and 3. Both hypotheses and findings will be discussed in the context of existing research on the neuroscience of intergroup biases.

Social Processing of Intergroup Affiliation

Study 2 examined how a sample of White perceivers formed impressions of happy smiling couples who were comprised of people who varied in racial composition. Confirmatory analysis in Study 2 centered on brain regions thought to support positive evaluations (i.e., NACC, OFC) and salience/ aversion (AI, amygdala). Additionally, exploratory analyses focused on the DACC, DLPFC, and VLPFC, regions thought to support the regulation of intergroup biases. In Study 2, we initially had two tentative confirmatory hypotheses for the patterns of activity in the OFC and NACC while perceivers processed images of couples. In the first, we predicted greater activity in either region to depictions of ingroup (i.e., perceived White couples) relative to the other conditions in light of work that suggests pro-ingroup reward biases (e.g., Cassidy & Krendl, 2016; Cikara et al., 2011; Hackel et al., 2015, 2020; Stanley et al., 2012; Wang et al., 2019; Willis et al., 2010). In the second, we predicted that if there were differences in positive evaluations between perceived Black and Interracial couples, there would be relatively greater brain activity in the NACC and OFC to depictions of perceived Black relative to depictions of Interracial couples, based on previous work that suggests that even among the general population, appraisals of Interracial relationships are appraised less positively than those of same-race couples by White perceivers (Skinner & Hudac, 2017; Skinner & Rae, 2019). However, results from Study 2 did not find these patterns and revealed almost the opposite of the anticipated direction of the results. That is, while perceiver's activity in the NACC and OFC did not significantly differ between perceived Black and perceived White couples, they

did show enhanced activity toward Interracial couples relative to either perceived Black or perceived White couples. There are two possibilities that may explain the unexpected pattern of results. The first may have been due to the content of the images displayed. Recall the OFC and NACC are thought to preferentially activate toward depictions of ingroup members, particularly when one's own explicit biases are salient or perceived outgroup members conflict with ingroup members (Cikara et al., 2011; Cikara & Fiske, 2012). Given the nature of the scenes depicting happy, smiling couples, it is highly unlikely that such depictions would signal intergroup conflict for a sample of White American college students. Furthermore, while it is possible that such depictions may interact with existing biases that participants may hold, this possibility seems increasingly unlikely given the increased exposure to depictions of Interracial relationships in media in recent years (Fitzpatrick, 2015; Mullins, 2014; Skinner & Rae, 2019) as well as approval for Interracial relationships being at historically high levels in the United States (McCarthy, 2021). In summary, while there was no difference in positive evaluations across either same-race couples, participants elicited greater activity in the NACC and OFC, which suggests that they may have evaluated the images of the Interracial couples more positively than the Black or White couples.

Similar to the regions supporting positive social evaluation, we offered two diverging confirmatory hypotheses for the amygdala and AI supporting social salience detection and aversion. We first predicted that despite Interracial romantic relationships becoming more commonplace in the United States (Fitzpatrick, 2015; Mullins, 2014; Pandey, 2022; K. Parker & Barroso, 2021), they are still a numeric

minority of relationships (Pew Research Center, 2017) and, in turn, may be more novel to perceivers than same-race romantic couples. Therefore, we first predicted that the amygdala and particularly the AI would track greater salience to the Interracial relative to either perceived Black or perceived White couples. Our second hypothesis was that if there were significant differences in activity in the AI but not the amygdala, this may indicate that the AI may be tracking more with aversive evaluations based on previous work concerning the role of the AI in social evaluation (Berntson et al., 2011; Borg et al., 2008; Calder et al., 2007; Heining et al., 2003; Simmons et al., 2004; Skinner & Hudac, 2017; Wicker et al., 2003). Results found increased activity toward Interracial couples relative to either perceived Black or perceived White couples in both the amygdala and AI, suggesting that participants may have found the depictions of Interracial couples more salient than the same-race couples. Because of the observed greater activity in both the amygdala and AI to depictions of Interracial couples relative to either Black or White couples, it seems that greater salience rather than evaluations may have been tracked by both regions. The findings in the amygdala and AI may be supported by previous work characterizing the role of the AI and amygdala in salience detection (Chekroud et al., 2014; Lau et al., 2020; Menon & Uddin, 2010; Uddin, 2015; Zhang & Peng, 2023).

While these patterns of results in the NACC and OFC may seem at odds with previous work that characterized social evaluations of Interracial couples as negative (Skinner & Hudac, 2017), the greater activity in the AI to Interracial couples did replicate the previously observed findings. Considering the results from the

confirmatory analysis in Study 2, the interpretation of Interracial couples being more salient may be the correct one for two primary reasons. First, it is difficult, if not impossible, to ignore the paradox that is created by greater activity being observed in the OFC and NACC to Interracial couples relative to Black and White couples if the increased AI activity is tracking negative evaluations. Based on these patterns, it does seem incredibly likely that this sample of perceivers did evaluate the more novel Interracial couples as more positive than same-race couples. Second, while Skinner & Hudac (2017) framed their results as tracking negative evaluations, their own findings do suggest the possibility that Interracial couples were viewed as more salient to their sample of participants rather than being viewed as more aversive. Their measure of aversion may have also been tracking salience, as the P300 ERP component has also been linked to motivational salience and is also highly context-dependent (Bartholow et al., 2006; Correll et al., 2006; Duval et al., 2013; Ito & Urland, 2003, 2005; Kubota & Ito, 2007; Kutas et al., 1977). To that end, it does appear that the heightened salience of Interracial couples is what was being tracked by greater AI and amygdala activity rather than more aversive or negative evaluations of the couples.

We similarly offered exploratory hypotheses for the brain regions supporting regulation and executive function (Carter et al., 1998; Carter & van Veen, 2007; Egner & Hirsch, 2005; MacDonald et al., 2000; Mansouri et al., 2009; Mattan, Wei, et al., 2018; Miller & Cohen, 2001; Ridderinkhof et al., 2004; Roberts & Hall, 2008). While we did not explicitly offer hypotheses for the VLPFC, we did offer exploratory hypotheses for the DLPFC and DACC. Within the context of viewing affiliative social

interactions, we were interested in the DACC in light of its role in conflict monitoring for intergroup biases among perceivers (Botvinick et al., 2001; Kolling et al., 2016). We were also interested in the DLPFC due to its role in executive function related to the selection of non-biased responses among perceivers when a conflict between intergroup biases and egalitarian beliefs is detected among perceivers (Badre & Wagner, 2004; Cassidy et al., 2016; Hartley & Phelps, 2010; Richeson et al., 2003; Sharp et al., 2004; Sreenivasan et al., 2014). We did not have strong hypotheses for the VLPFC as it has been inconsistently implicated in intergroup bias regulation (c.f., Telzer et al., 2015). Furthermore, we were particularly interested in the DLPFC and DACC because both regions are believed to be related to the detection of conflict between egalitarian worldviews and responses among people who are highly motivated to regulate their responses (Amodio, 2014; Amodio et al., 2004, 2006, 2008; Ito & Bartholow, 2009; Kubota et al., 2012), particularly those who are strongly externally motived to do so (Amodio et al., 2004, 2006, 2008). Because the acceptance of Interracial dating is more normative in the US relative to any point in US history (McCarthy, 2021; Pandey, 2022; K. Parker & Barroso, 2021), it was possible that even if some perceivers did not explicitly approve of Interracial dating or marriage, they might be more likely to try to regulate their responses in line with societal norms. Therefore, we predicted that this sample of White American perceivers would show increased activity in both regions to depictions of Interracial dating relative to either condition of same-race couples. Further, we initially predicted these patterns of activity considering work characterizing depictions of Interracial dating as being

evaluated less positively than depictions of dating when both partners were perceived to be of the same race (Skinner & Hudac, 2017).

Our initial predictions for the DACC and DLPFC were correct as in both regions; we found greater activity toward the depictions of Interracial couples relative to either perceived Black or White couples. Additionally, we found greater activity in the VLPFC to depictions of Interracial couples relative to either perceived Black or White couples. Considering the patterns of results from the regions supporting positive evaluations and salience, we offer an updated interpretation of the results observed in the DLPFC, DACC, and VLPFC. Specifically, it appears that it was because Interracial couples were perceived as being more salient than perceived Black or White couples; participants may have preferentially engaged in executive function and response monitoring in the interest of detecting conflicts between their responses and their own worldviews. Taken together with the results from the regions supporting social evaluation and salience detection, it seems possible that participants may have effortfully engaged regulatory mechanisms when viewing the more salient depictions of Interracial couples. However, they may have done so in the interest of evaluating these depictions more positively than either condition of same-race dating. If this interpretation is correct, it does support previous accounts of the DACC and DLPFC working in concert to monitor responses and respond in an egalitarian manner (Amodio, 2014; Amodio & Cikara, 2021; Kubota et al., 2012; Mattan, Wei, et al., 2018). In short, to make sense of relatively more salient stimuli and ultimately form positive social evaluations of depictions of Interracial couples, participants may have

needed to actively engage in response monitoring to more effectively perceive and evaluate the depictions of Interracial dating.

In addition to the discussion of the observed results from Study 2 in the previous paragraphs, in the following paragraphs, we discuss some additional factors that we believe may have contributed to the observed confirmatory fMRI findings and, in turn, speculate about how our findings may have changed due to these variables. First, our sample of participants was generally younger college-aged students (M_{age} = 19.143, SD_{age} = 2.336). While we do not speculate there would be large differences in the patterns of results observed in Study 2 as a function of participant age, there are some reasons to consider the possibility of participant age moderating the effects obtained in regions supporting positive evaluations, salience, and bias regulation. First, while approval of Interracial dating and marriage is currently at historically high levels, polling data in the United States suggests that the majority of people in the US did not approve of Interracial relationships until 1997 (Pew Research Center, 2017). Given the age of participants in the study, this shift occurred nearly 10 years before they were born. Therefore, it may be possible that the younger participants viewed Interracial dating as more normative and may have been less exposed to negative attitudes or associations of Interracial dating and romance. If the sample was also comprised of older people as well, it is possible that the older participants may still have, at the very least, less positive associations of Interracial relationships even if they explicitly approved of them. As such, it is possible that perhaps the older participants would show relatively greater engagement of the amygdala and AI and

diminished OFC and NACC activity to the Interracial couples relative to perceived Black or White couples.

In addition to participant age potentially influencing the observed results, it is possible that the gender composition of the sample may have impacted the results, specifically with respect to the degree to which participants endorsed sexism. Recall the majority of the sample consisted of young self-identified adult women (39 Women, 9 Men, 1 Genderfluid). This raises the possibility that the results from Study 2 may have been moderated by the explicit endorsement of sexism if the sample of the study consisted of only young adult cisgender men. Given concerning trends of greater emergence or acceptance of misogyny among college-age cisgender men in the past decade (Ging, 2019; Marwick & Caplan, 2018; O'Hanlon et al., 2024; Over et al., 2025; Perliger et al., 2022) it is possible that having a sample comprised of younger cisgender men may have shown different evaluative responses to depictions of couples than a sample of adult cisgender women may have. Considering the link between the so-called "manosphere" and men's rights advocates with white supremacist and racially prejudiced worldviews (McMahon & Kahn, 2018; Over et al., 2025; Ribeiro et al., 2021), it is then possible that sexism may have also impacted how participants would have evaluated the couples. For example, greater endorsement of sexism among a sample of men may be associated with diminished NACC and OFC activity toward Interracial couples relative to White couples. While we do not want to speculate too much on how differences in sexism endorsement may have changed our results, we

acknowledge that future investigations may be further contextualized by measuring participants' endorsement of misogynistic worldviews.

Another variable that may have changed our results was whether participants were actively in a relationship at the time of the study, as well as how happy with their current relationship status they may be. For example, it is tentatively possible that if participants were not actively in a relationship or recently ended a relationship, the affiliative nature of the couples photos may be interpreted differently than if they actively were in a relationship. If they were dissatisfied about not being in a relationship, they may possibly show diminished activity in the NACC or OFC and possibly even greater AI activity to the images of couples relative to those who were not in a relationship. One interesting possibility is if a dissatisfaction with not being in a relationship could create a perception of conflict or scarcity for those participants. If this was the case, one might expect that activity in the AI may reflect greater aversion to outgroup members relative to ingroup members, similar to previous work on intergroup dynamics (e.g., De Dreu, 2010; Halevy et al., 2008, 2012; Krosch et al., 2017; Krosch & Amodio, 2019).

Lastly, another variable that may have changed the observed results was if the depictions of affiliative stimuli were framed in the first person relative to the third person. That is, participants processed images of couples where they were a third-person observer rather than directly interacting with a prospective romantic partner. If participants, however, viewed images of perceived Black and White potential dating partners with a cover story suggesting this was someone they could be in a

relationship with, it is possible that we may have seen patterns of NACC, OFC, amygdala, and AI activity in line with a preference for ingroup members. Specifically, we may have seen greater OFC and NACC activity to perceived White partners relative to perceived Black partners (e.g., Amodio, 2014; Firat et al., 2017; Hackel et al., 2017, 2020; Van Bavel et al., 2008) as well as greater AI and amygdala activity to perceived Black partners relative to perceived White partners (Krill & Platek, 2009; Merritt et al., 2021; Shkurko, 2013; Stanley et al., 2012; Van Bavel et al., 2008).

In summary, our confirmatory results suggest that this sample of White American perceivers may have viewed depictions of Interracial couples more positively than depictions of either perceived Black or White couples. Furthermore, these results may suggest that this sample of participants potentially viewed depictions of Interracial couples as also being more salient compared to the depictions of perceived Black or White couples. Finally, the results from our exploratory analysis may suggest that participants engaged in more deliberate regulation of the responses toward Interracial couples relative to perceived Black or White couples, and the greater engagement of regulation may have possibly been in service of effectively evaluating the Interracial couples relative to perceived Black or White couples.

Social Processing of Intergroup Violence

Study 3 examined how a sample of White perceivers formed impressions of either mundane social interactions or depictions of violence where the victim was either perceived as Black or White. Confirmatory analysis in Study 3 centered on brain

regions thought to support salience/ aversion (AI, amygdala) and mentalizing (TPJ, STS). Additionally, exploratory analyses focused on the DMPFC, a region thought to support mentalizing, and the DACC, DLPFC, and VLPFC, regions thought to support the regulation of intergroup biases. In Study 3, we initially predicted two tentative confirmatory hypotheses for patterns of activity in the AI and amygdala when participants viewed the violent images depending on whether either the AI or amygdala were tracking salience or evaluative responses. Similar to the predictions offered in Study 2 for the AI and amygdala, the tentative hypotheses were dependent on whether activity in either region supported salience detection or aversion.

Results in the AI between the two violent conditions seem to suggest that, at least among the violent images, this sample of participants may have found depictions of perceived Black victims to be more aversive than depictions of perceived White victims. While we had initially predicted that if there was a difference in AI activity between violent image conditions, AI activity would be greater toward perceived White victims relative to perceived Black victims. We predicted this due to previous literature that found perceivers showed greater activity in the AI when members of ingroups experienced a negative outcome relative to when members of outgroups experienced the same outcome (Cikara et al., 2011; but see Merritt et al., 2021; Shkurko, 2013). However, our results demonstrated the opposite, as there was greater AI activity toward the depictions of perceived Black victims relative to the depictions of White victims. These findings may be explained by existing work that has found that pro-ingroup biases may be dampened or reversed when an outgroup is harmed,

and such harm is viewed as unfair or unwanted (Mathur et al., 2010). Within the context of the stimuli, these patterns of AI activity may also be influenced by members of law enforcement or the military being the aggressors in the majority of the violent images. Considering the pervasiveness of depictions of law enforcement brutality against perceived Black Americans (Campbell & Valera, 2020; Canevez et al., 2022), it is possible that these depictions, in particular, were more aversive, or at the very least viewed as less justified or fair by perceivers. As such, they may not have shown a pro-ingroup bias between violent image conditions because of the pervasive and negative association of police brutality toward perceived Black Americans (Campbell & Valera, 2020; Canevez et al., 2022).

If either region supported salience (Chekroud et al., 2014; Uddin, 2015), we initially hypothesized that there would be greater activity in both regions to depictions of violence relative to the neutral social interactions. Our results did not support this hypothesis, as there was no significant difference in amygdala activity across conditions. However, there was greater activity in the AI to neutral scenes relative to violent scenes. While unexpected, these patterns of results may be explained by greater ambiguity of the behaviors and mental states of the individuals depicted in the neutral scenes relative to the violent ones. The violent scenes were more consistent with respect to the behaviors depicted in each scene, whereas there was greater variability in the behaviors and facial expressions for the neutral scenes. To this end, it is possible that the relative uniformity of the behaviors and emotions displayed in the violent images may have potentially made the greater variety of behaviors on display

in the mundane social scenes appear to be relatively more novel than the violent ones. If this was what was driving the differences in AI activity between the neutral and violent scenes, then this may tentatively support previous findings in the AI that appear to reflect processing salience due to perceived novelty (Cloutier et al., 2017; Cloutier, Li, et al., 2014; Lau et al., 2020).

We additionally offered two confirmatory hypotheses on regions previously demonstrated to support mentalizing. We first predicted that if there were differences in activity among regions supporting mentalizing between the two violent image conditions, there would be greater activity toward the depictions of perceived Black victims relative to those that depicted the perceived White victims. While this prediction does not necessarily reflect pro-ingroup bias, it was predicted in light of more recent work that found greater activity in the TPJ and STS toward outgroup members relative to ingroup members when perceivers inferred the mental states of others (Dang et al., 2022; Handley et al., 2022, 2023). The results of Study 3 did partially support this hypothesis, specifically in that there was greater TPJ activity to depictions of perceived Black victims, in line with existing work that found similar patterns of mentalizing in the TPJ toward depictions of perceived Black people being forcefully arrested relative to depictions of perceived White people being forcefully arrested (Dang et al., 2022). As such, these results are generally aligned with previous work that found more effortful mentalizing toward depictions of outgroup members being treated violently by White aggressors.

We also predicted that there would be greater activity in the STS and TPJ toward violent scenes relative to the neutral ones, considering work that found depictions of violence were associated with greater activity in the STS and TPJ (Bruneau et al., 2012; Dang et al., 2022; Young et al., 2011). Results, however, revealed this to not be the case. There was no significant difference in TPJ activity between violent or neutral scenes, but there was greater activity in the STS toward the neutral scenes relative to either violent condition. Similarly to the findings in the amygdala and AI, this may be again a result of the relative visual ambiguity of the neutral stimuli relative to the violent scenes. Because the STS is implicated in mentalizing and processing complex visual stimuli (Cloutier, Turk, et al., 2008; Dang et al., 2022; Deen et al., 2015; Handley et al., 2022, 2023; Hoffman & Haxby, 2000; Schurz et al., 2014) this may explain the patterns of activity between the violent and non-violent stimuli conditions in the STS. This would mean that in the effort to better understand and infer the thoughts and intentions of individuals in the neutral scenes, more engagement of the STS was necessary to effectively process the neutral scenes relative to the violent ones because of the relatively more ambiguous nature of the social information depicted in the violent scenes.

Lastly, for the exploratory analyses with the DACC, DLPFC, and VLPFC, we hypothesized that there would be greater activity in all three regions to the depictions of violence relative to the neutral scenes. We predicted this pattern of results in light of the aversive nature of the images and that greater activity in these three regions would be related to downregulating the unpleasant nature of the violent images

(Denny et al., 2023; Hartley & Phelps, 2010; Ochsner et al., 2012; Tupak et al., 2014; Wager et al., 2008) or in support of more deliberate processing of the images in efforts to align evaluations and mentalizing with egalitarian worldviews under the assumption that both violent conditions were equally as aversive (Amodio, 2010, 2014; Amodio et al., 2006, 2008; Amodio & Cikara, 2021; Cassidy et al., 2016; Richeson et al., 2003; Telzer et al., 2015). Results first found greater activity in the DLPFC to depictions of perceived Black victims relative to perceived White ones. The greater engagement of the DLPFC was in line with our initial confirmatory predictions in that we predicted increased DLPFC activity to perceived Black victims in the service of participants monitoring their own responses when processing the violent scenes (Amodio, 2014; Cassidy et al., 2016; Kubota et al., 2012; Richeson et al., 2003). Considering the patterns of results for brain regions supporting mentalizing and negative evaluations of the scenes, it seems possible that greater DLPFC activity may have been required for this sample of participants to more actively engage response monitoring while viewing these images (Cassidy et al., 2016; Sreenivasan et al., 2014), and in doing so may have been able to evaluate the depictions of perceived Black victims as more aversive relative to perceived White victims (Zhang & Peng, 2023). Moreover, results found greater activity in all three of these regions toward the depictions of neutral social interactions relative to either violent condition. While it is not immediately clear why this pattern of results emerged, it may tentatively be attributed to processing the relative ambiguity of the emotions on display from individuals depicted in the neutral social scenes compared to those depicted in the violent scenes (Sun et al., 2023).

Similar to the discussion of Study 2, in the following paragraphs, we offer a brief discussion on some factors that may have changed our patterns of results and speculate as to how our observed patterns of results may have changed if these variables were different. First, the homogeneity of the age of participants in our sample may explain our current results for two reasons. One reason that participants' age may have influenced the observed results may lie with the relative exposure to violence in media (Huesmann, 2007) or in the real world (Mrug et al., 2016). Given the relative prevalence of violence in the zeitgeist, it is possible that participants may be desensitized to depictions of violence (Mrug et al., 2016). While it is unclear whether older participants would be similarly desensitized to violence relative to the younger participants, the inclusion of older participants may highlight differences in the processing of the violent scenes due to perceptions of lower social distance. That is, because the violent stimuli were rated as being believed to take place in the 1980s-1990s, the images may have been less relevant to the younger participants due to greater social distance from the events (Mather et al., 2017). In either case, if the age of the participants was associated with evaluations of the stimuli, then it may be possible that while participants may still show greater AI activity to Black victims, the older participants may have also shown greater AI activity to the violent relative to neutral images.

We cautiously acknowledge that our sample, which was primarily comprised of self-identified cisgender women, may have influenced our results. However, based on the observed results, we do not believe that our findings in Study 3 may be

attributable to the participant's gender identity. We would not have predicted that a sample comprised solely of either cisgender men or cisgender women would have reflected differences in processing violence between Black and White victims. If differences in processing of the violent stimuli did arise, then we tentatively may have expected to see a reduction in aversion to the violent images relative to the neutral ones for a sample comprised only of cisgender men due to greater exposure to physical violence among men and boys (Finkelhor et al., 2009; Scott-Storey et al., 2018). However, because in our results, we found greater activity in the AI to the neutral scenes relative to the ones depicting perceived White victims, the diminished aversion as a function of participant gender does seem highly unlikely.

Lastly, we speculate as to how our observed results may have differed if the images were framed within the first person instead of the third person, with the participant being the victim rather than a third person. First, we strongly believe that if the participants were framed as the victim of violence, then both the amygdala and AI would show significantly greater activity to the violent scenes relative to the neutral ones (Borg et al., 2008; Hariri et al., 2002; Öhman, 2005). Additionally, we believe that the perceived threat posed by the images of a perceived White or Black aggressor would lead to more salient anti-outgroup biases. Specifically, based on previous work showing greater negative associations of threatening outgroup members, we would anticipate seeing greater activity in both the AI and amygdala to perceived Black relative to perceived White aggressors (Domínguez D et al., 2018; Krill & Platek, 2009; Molapour et al., 2015; Phelps et al., 2000).

In summary, our confirmatory results suggest that this sample of White American perceivers may have viewed depictions of violence toward Black victims as more aversive than violence toward White victims. The results in the AI also may suggest that participants found the neutral scenes to be more salient relative to the violent images. Furthermore, our results in the TPJ suggest that among the violent images, participants more effortfully tried to infer the thoughts and intentions of the Black victims relative to the White victims. Our results in the STS also may suggest that participants needed to exert more effort to effectively infer mental states from the more ambiguous nature of the emotions and behaviors in the neutral scenes relative to the violent ones. Finally, our exploratory results in the DLPFC suggest that participants are more effortfully engaged in response monitoring when processing the images of Black victims relative to the White victims, possibly in support of greater efforts to mentalize with the Black victims.

Theoretical Implications

The findings from this dissertation add valuable perspective to theory within social psychology. Foundational work within the field has characterized affinity for one's ingroup, or ingroup love (Allport, 1954; Brewer, 1999; Tajfel & Turner, 1986), along with degradation of outgroup members, as being strongly related to the formation of intergroup biases (De Dreu, 2010; Halevy et al., 2008, 2012; M. T. Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). Within this Social Identity Theory, it is thought that people will generally favor members of their perceived

ingroups and will act less favorably to those who are members of outgroups. Previous efforts have demonstrated that people may show preferences toward their ingroups and may in turn view outgroup members negatively (Amira et al., 2021; Brewer, 1999; Halevy et al., 2012; Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015), or even act in a discriminatory manner (Allport, 1954; Amira et al., 2021; Lelkes & Westwood, 2017).

This dissertation, however, suggests that the relationship between social identity and intergroup biases may be more complex than previously thought. (Amodio, 2014; Amodio et al., 2006, 2007, 2008; Amodio & Cikara, 2021; Plant & Devine, 1998). In Studies 2 and 3 of this dissertation, participants showed diminished pro-ingroup bias in evaluative and cognitive mechanisms when processing relatively complex social stimuli. It is possible that pro-ingroup preferences may have been present if the sample consisted of individuals who had some degree of intergroup bias, particularly if either of the affiliative or aversive stimuli indicated some degree of scarcity or conflict. Scarcity or perceived conflict is thought to increase more negative evaluations or associations of outgroup members (Bruneau et al., 2012; Bruneau & Saxe, 2010; Cikara, 2015; Cikara & Fiske, 2012, 2013; Krosch et al., 2017). To that end, future expansions to work centered on Social Identity Theory (Allport, 1954; Tajfel & Turner, 1986) should also consider how explicitly held biases may further interact with the processing of affiliative or aversive social interactions.

This dissertation also has implications for future work concerning the formation of extremely biased worldviews, including the formation of extreme white supremacist

worldviews. The two sets of stimuli are uniquely suited to examine how individuals who hold white supremacist or extremist views may differentially process complex social interactions across individuals with different racial compositions. White supremacist worldviews include increased explicit prejudicial attitudes towards interracial dating and marriage (Stillwell & Lowery, 2021). Specifically, individuals who increasingly adopt white supremacist extremist worldviews, in particular, tend to have extremely negative evaluations of the mere idea of interracial marriage and dating (Simi et al., 2017). Indeed, negative evaluations of interracial dating and marriage are seen as visceral betrayals of ingroup values to members of white supremacist extremist groups (Castle et al., 2020; Simi et al., 2017).

Aversion to interracial dating and marriage frequently constitutes elements of pervasive rallying cries among extremists. Specifically, strong endorsement of the great replacement theory (i.e., the conspiracy theory that white individuals in the United States and Western Europe will be replaced or assimilated by non-white individuals; Ong, 2020) has previously been associated with actions of terrorism at the hands of white supremacist extremists (Ong, 2020; Pape, 2021). For these individuals, depictions of interracial marriage or dating are not only seen as a violation of a moral imperative but also as an actual threat to the "white race" and "whiteness" (Ong, 2020) and, therefore, are leveraged heavily in white supremacist propaganda. As such, the stimuli curated in Study 1A may be uniquely suited to probe how the processing of intergroup affiliation differs among perceivers who endorse white supremacist worldviews relative to those who do not. Comparisons of the neural correlates of

intergroup processing across groups of people who differentially endorse white supremacist worldviews may further elucidate mechanisms that may contribute to the further endorsement of white supremacist worldviews.

In addition to the affiliative social interactions, the depictions of aversive social interactions curated in Study 1B may similarly be leveraged to examine the neural correlates of extremist ideologies. Within white supremacist propaganda, violence is often portrayed to show devotion to a broader cause (Beck, 2012; Bracke & Hernández Aguilar, 2020; Glaser et al., 2002; Kruglanski, 2018; Kruglanski et al., 2014, 2018; McVeighn, 1999; Quillian, 1995; Rousis et al., 2020; Van Dyke & Soule, 2002; Webber & Kruglanski, 2018). The positive portrayal of violence is also reinforced as a value within families (Simi, Futrell, et al., 2016; Simi, Sporer, et al., 2016; Windisch et al., 2020). For example, former white supremacists report instilling violent rhetoric within their families, with adult members framing violence as a positive action to their children from a young age (Simi, Futrell, et al., 2016; Windisch et al., 2020). Violent rhetoric also provides a means to sustain white supremacist groups[10] (Simi, Futrell, et al., 2016). As such, the depictions of violence may be leveraged to examine how greater endorsement of white supremacist worldviews may

[10] Ironically, the overt emphasis on violence may, in some cases, prevent a larger-scale mobilization of these groups' efforts, as membership within white supremacist extremist groups may not always be stable. Accounts for former right-wing extremists have reported infighting within the groups as strongly contributing to their deradicalization and exit from the group (Windisch et al., 2019).

moderate the activity of brain regions that support the processing of aversive social interactions. More specifically, characterizing the differences in activity across brain networks as a function of the endorsement of white supremacy may provide more information about the affective and cognitive correlates that may explain further entrenchment in extremist ideologies.

To that end, the depictions of Interracial dating and violence can be used in future work to examine the difference in engagement of evaluative and cognitive mechanisms between those with extremist worldviews, those who only have explicitly held biases, and those who do not. Quantifying the differences across brain regions that support affect and cognition may further explain the escalation of bias into extremist ideologies and, more importantly, may ultimately lead to the development of interventions to curb entrance into white supremacist extremist movements.

Limitations and Future Directions

In this section, we briefly summarize the limitations of the studies detailed in this dissertation. Moreover, we offer some future directions for subsequent studies concerning the social processing of affiliative and aversive social interactions. First, the overall generalizability of these findings is somewhat limited to self-identified White American undergraduate students, as the sample tended to be comprised of relatively younger college students. Second, while the stimuli that were piloted in Study 1B were rigorously piloted and equated, the relative variability of behaviors and emotions displayed in the neutral images may have made them more difficult to

process than the violent ones. This difference between images may have been a potential confound leading to greater engagement of salience, mentalizing, and regulation mechanisms when viewing the non-violent scenes relative to the violent ones. Relatedly, all images depicting cross-race social interactions were comprised of individuals who were perceived to be Black or White, limiting the generalizability of the findings of either study to groups of people who may not be perceived as either Black or White.

Next, while both Study 2 and Study 3 found engagement of regulatory mechanisms that may be explained by individual differences in motivation to regulate prejudice, it is difficult to make this claim with absolute certainty in the absence of a measure of their expressed motivation to regulate prejudice for the entire sample (Plant & Devine, 1998). However, a future direction that may better contextualize the results from the exploratory analysis on regions supporting regulation would be to see if IMS/EMS did moderate activity in the DACC, DLPFC, and VLPFC. This would be feasible as a large number of the sample did respond to these measures as part of their participation in the University of Delaware's PSYC 100 Sona pool. Similarly, the inclusion of a measure of intergroup contact for the entire sample may have provided more clarity to the results that found greater reward and salience to the depictions of Interracial dating relative to those depicting either Black or perceived White couples. However, it may be similarly possible to examine differences in interracial contact for a number of the participants as they also completed a measure of interracial contact as part of the University of Delaware's PSYC 100 Sona pool. Furthermore, both Studies

2 and 3 did not focus on the endorsement of explicit racial bias. However, future investigations that build upon either study should strongly consider examining the role of explicit racial bias, or the endorsement of white supremacist worldviews may moderate the social processing of either intergroup harmony or violence because of the relevance to propaganda used by white supremacists (e.g., Castle et al., 2020; Kruglanski, 2018; Kruglanski et al., 2014, 2018; Ong, 2020; Simi et al., 2017; Simi, Futrell, et al., 2016; Simi, Sporer, et al., 2016; Webber & Kruglanski, 2018). This next step may also serve as a natural springboard for other investigations, which may examine how both current and former white supremacist extremists may differentially process these scenes, given the relevance of either dimension of stimuli to their explicit worldviews (Glaser et al., 2002; Kruglanski, 2018; Rousis et al., 2020; Simi et al., 2017; Simi, Futrell, et al., 2016; Simi, Sporer, et al., 2016; Webber & Kruglanski, 2018). Such comparisons may help inform new frameworks that can better account for the affective, behavioral, and cognitive mechanisms that serve to drive individuals who endorse explicit biases to those who outright endorse extremist white supremacist worldviews and may ultimately become radicalized into white supremacist extremist groups.

General Conclusion

This dissertation adds to our understanding of how people may process more complex depictions of intergroup harmony and conflict. Drawing on previous work from psychology and neuroscience, this dissertation first led to the creation of two

stimuli sets, both visual representations of intergroup affiliation and aversive depictions of cross-race violence. Perceivers showed greater activation of brain regions underlying reward and positive evaluations of depictions of Interracial dating relative to those depicting same-race couples. Further, in light of greater activity to depictions of Interracial couples relative to perceived Black or White couples in regions supporting reward, we demonstrated that perceivers might not view depictions of Interracial dating as aversive but rather novel and, therefore, salient. In addition, this dissertation highlights how perceivers may more negatively evaluate depictions of violence toward outgroup members. Furthermore, this thesis replicated previous work on intergroup mentalizing as it demonstrated that perceivers will try to infer the thoughts and intentions of outgroup members who are being treated violently preferentially to ingroup members who are being treated violently. Taken together, the findings from Studies 2 and 3 suggest that among White perceivers, there may not always be a tendency to only preferentially evaluate or mentalize with members of their ingroup. Rather, these findings suggest that White perceivers may, in some cases, show preferential evaluations and mentalize toward outgroup members when viewing more complex depictions of affiliative or aversive social behaviors.

As such, this dissertation establishes a baseline for how White American perceivers may process more complex intergroup social interactions and highlights the utility of examining intergroup dynamics by using depictions of social behaviors rather than the more conventional use of face stimuli. This thesis may also serve as a basis to better understand the neural correlates that may contribute to further

entrenchment in explicit intergroup biases that lead to the endorsement of extremist worldviews and the adoption of extremist identities. In a world that is rapidly seeing the rise of far-right, prejudiced populism, these future directions of work could not be timelier. It is our hope that this dissertation will serve as a starting point for future research that will help aid interventions to the escalation of such ideologies among those who are most at risk of developing them.

REFERENCES

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., Manes, F., & Ibáñez, A. (2017). Convergence of interoception, emotion, and social cognition: a twofold fMRI meta-analysis and lesion approach. *Cortex*, *88*, 124–142.

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annu. Rev. Psychol.*, *60*(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Allport, G. (1954). *The nature of prejudice*. Addison-Wesley.

Ames, D. L., Jenkins, A. C., Banaji, M. R., & Mitchell, J. P. (2007). Neural correlates of automatic beliefs about gender and race. *Human Brain Mapping*, *28*(10), 915–930. https://doi.org/10.1002/hbm.20320

Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2021). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, *43*, 473–494.

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition*, *28*(6), 695–716.

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3800

Amodio, D. M., & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, *72*(1), 439–469.

Amodio, D. M., Devine, P. G., & Harmon-jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, *18*(6), 524–531.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*, *94*, 60–74.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*(2), 88–93. http://pss.sagepub.com/content/15/2/88.abstract

Amodio, D. M., Kubota, J. T., Harmon-jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs. external cues. *Social Cognitive and Affective Neuroscience*, *1*(1), 26–36. https://doi.org/10.1093/scan/nsl002

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185.

Aron, A. R., Robbins, T. W., Poldrack, R. A., AR, A., TW, R., RA, P., Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170. https://doi.org/10.1016/j.tics.2004.02.010

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology : CB*, *20*(11), 1018–1022. https://doi.org/10.1016/j.cub.2010.03.071

Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, *41*(3), 473–487.

Barth, D. M., Mattan, B. D., Dang, T. P., & Cloutier, J. (2020). Regional and network neural activity reflect men's preference for greater socioeconomic status during impression formation. *Scientific Reports*. https://doi.org/10.1038/s41598-020-76847-z

Bartholow, B. D., & Dickter, C. L. (2007). *Social neuroscience: Integrating biological and psychological explanations of social behavior* (E. Harmon-Jones & P. Winkielman, Eds.; Issue 573). Guilford Press.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: cognitive control of inhibition and its impairment by alcohol. *J Pers Soc Psychol*, *90*(2), 272–287. https://doi.org/10.1037/0022-3514.90.2.272 [doi]

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, *12*(3), 197–204. https://doi.org/10.1111/1467-9280.00336

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

Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, *33*(6), 1452–1469.

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R. R., & Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive and Affective Neuroscience*, *9*(5), 653–660.

Beaver, J. D., Lawrence, A. D., Van Ditzhuijzen, J., Davis, M. H., Woods, A., & Calder, A. J. (2006). Individual differences in reward drive predict neural responses to images of food. *Journal of Neuroscience*, *26*(19), 5160–5166.

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2008). The Quadruple Process model approach to examining the neural underpinnings of prejudice. *NeuroImage*, *43*(4), 775–783. https://doi.org/10.1016/J.NEUROIMAGE.2008.08.033

Berkman, E. T., Cunningham, W. A., & Lieberman, M. D. (2014). Research methods in social and affective neuroscience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (1st ed., pp. 123–158). Cambridge University Press New York.

Berntson, G. G., Norman, G. J., Bechara, A., Bruss, J., Tranel, D., & Cacioppo, J. T. (2011). The insula and evaluative processes. *Psychological Science*, *22*(1), 80–86.

Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*, *199*(3), 457–480. https://doi.org/10.1007/s00213-008-1099-6

Borg, J. S., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, *20*(9). https://doi.org/10.1162/jocn.2008.20109

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

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. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B. (2017). *Intergroup discrimination: Ingroup love or outgroup hate?*

Broyd, S. J., Richards, H. J., Helps, S. K., Chronaki, G., Bamford, S., & Sonuga-Barke, E. J. S. (2012). An electrophysiological monetary incentive delay (e-MID) task: a way to decompose the different components of neural response to positive and negative monetary reinforcement. *Journal of Neuroscience Methods*, *209*(1), 40–49.

Bruneau, E. G., Dufour, N., & Saxe, R. (2012). Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 717–730.

Bruneau, E. G., Jacoby, N., Kteily, N., & Saxe, R. (2018). Denying humanity: The distinct neural correlates of blatant dehumanization. *Journal of Experimental Psychology. General*, *147*(7), 1078–1093. https://doi.org/10.1037/xge0000417

Bruneau, E. G., & Saxe, R. (2010). Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *NeuroImage*, *52*(4), 1704–1711. https://doi.org/10.1016/j.neuroimage.2010.05.057

Burke, S. E., Dovidio, J. F., Perry, S. P., Burgess, D. J., Hardeman, R. R., Phelan, S. M., Cunningham, B. A., Yeazel, M. W., Przedworski, J. M., & van Ryn, M. (2017). Informal Training Experiences and Explicit Bias against African Americans among Medical Students. *Social Psychology Quarterly*, *80*(1), 65–84. https://doi.org/10.1177/0190272516668166

Butz, D. A., & Plant, E. A. (2009). Prejudice control and interracial relations: The role of motivation to respond without prejudice. *Journal of Personality*, *77*(5), 1311–1342.

Cacioppo, J. T., Amaral, D. G., Blanchard, J. J., Cameron, J. L., Carter, C. S., Crews, D., Fiske, S., Heatherton, T., Johnson, M. K., & Kozak, M. J. (2007). Social neuroscience: Progress and implications for mental health. *Perspectives on Psychological Science*, *2*(2), 99–123.

Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., & McClintock, M. K. (2000). Multilevel integrative analyses of human behavior: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin*, *126*(6), 829–843. https://doi.org/10.1037/0033-2909.126.6.829

Calder, A. J., Beaver, J. D., Davis, M. H., Van Ditzhuijzen, J., Keane, J., & Lawrence, A. D. (2007). Disgust sensitivity predicts the insula and pallidal response to pictures of disgusting foods. *European Journal of Neuroscience*, *25*(11), 3422–3428.

Campbell, F., & Valera, P. (2020). “The only thing new is the cameras”: A study of US college students’ perceptions of police violence on social media. *Journal of Black Studies*, *51*(7), 654–670.

Canevez, R. N., Karabelnik, M., & Winter, J. S. (2022). Police brutality and racial justice narratives through multi-narrative framing: Reporting and commenting on the george floyd murder on youtube. *Journalism & Mass Communication Quarterly*, *99*(3), 696–717.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 367–379. https://doi.org/10.3758/CABN.7.4.367

Cassidy, B. S., & Krendl, A. C. (2016). Dynamic neural mechanisms underlie race disparities in social cognition. *NeuroImage*, *132*, 238–246.

Cassidy, B. S., Lee, E. J., & Krendl, A. C. (2016). Age and executive ability impact the neural correlates of race perception. *Social Cognitive and Affective Neuroscience*, *11*(11), 1752–1761.

Castle, T., Kristiansen, L., & Shifflett, L. (2020). White racial activism and paper terrorism: a case study in far-right propaganda. *Deviant Behavior*, *41*(2), 252–267.

Chakrabarti, B., Haffey, A., Canzano, L., Taylor, C. P., & McSorley, E. (2017). Individual differences in responsivity to social rewards: Insights from two eye-tracking tasks. *Plos One*, *12*(10), e0185146.

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: a review and synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300.

Chekroud, A. M., Everett, J. A. C., Bridge, H., & Hewstone, M. (2014). A review of neuroimaging studies of race-related prejudice: does amygdala response reflect threat? *Frontiers in Human Neuroscience*, *8*, 179. https://doi.org/10.3389/fnhum.2014.00179

Childress, A. R., Ehrman, R. N., Wang, Z., Li, Y., Sciortino, N., Hakun, J., Jens, W., Suh, J., Listerud, J., & Marquez, K. (2008). Prelude to passion: limbic activation by “unseen” drug and sexual cues. *PLoS One*, *3*(1), e1506.

Cikara, M. (2015). Intergroup schadenfreude: Motivating participation in collective violence. *Current Opinion in Behavioral Sciences*, *3*, 12–17.

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.

Cikara, M., & Fiske, S. T. (2012). Stereotypes and schadenfreude: Affective and physiological markers of pleasure at outgroup misfortunes. *Social Psychological and Personality Science*, *3*, 63–71. https://doi.org/10.1177/1948550611409245

Cikara, M., & Fiske, S. T. (2013). Their pain, our pleasure: Stereotype content and schadenfreude. *Annals of the New York Academy of Sciences*, *1299*, 52–59. https://doi.org/10.1111/nyas.12179

Cikara, M., Van Bavel, J. J., Ingbretsen, Z. A., & Lau, T. (2017). Decoding "us" and "them": Neural representations of generalized group concepts. *Journal of Experimental Psychology. General*, *146*(5), 621–631. https://doi.org/10.1037/xge0000287

Cloutier, J., Ambady, N., Meagher, T., & Gabrieli, J. D. E. (2012). The neural substrates of person perception: Spontaneous use of financial and moral status knowledge. *Neuropsychologia*, *50*(9), 2371–2376. https://doi.org/10.1016/j.neuropsychologia.2012.06.010

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PloS One*, *9*(1), e84677.

Cloutier, J., Gabrieli, J. D. E., O'Young, D., & Ambady, N. (2011). An fMRI study of violations of social expectations: When people are not who we expect them to be. *NeuroImage*, *57*(2), 583–588. https://doi.org/10.1016/j.neuroimage.2011.04.051

Cloutier, J., & Gyurovski, I. (2014). Ventral medial prefrontal cortex and person evaluation: Forming impressions of others varying in financial and moral status. *Neuroimage*, *100*, 535–543. https://doi.org/10.1016/j.neuroimage.2014.06.024

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *Journal of Cognitive Neuroscience*, *20*, 941–951. https://doi.org/10.1162/jocn.2008.20062

Cloutier, J., Li, T., & Correll, J. (2014). The impact of childhood experience on amygdala response to perceptually familiar Black and White faces. *Journal of Cognitive Neuroscience*, *26*(9), 1992–2004. https://doi.org/10.1162/jocn_a_00605

Cloutier, J., Li, T., Mišić, B., Correll, J., Berman, M. G., Mišic, B., Correll, J., & Berman, M. G. (2017). Brain Network Activity During Face Perception: The Impact of Perceptual Familiarity and Individual Differences in Childhood Experience. *Cerebral Cortex*, *27*, 1–13. https://doi.org/10.1093/cercor/bhw232

Cloutier, J., Turk, D. J., & Macrae, C. N. (2008). Extracting variant and invariant information from faces: The neural substrates of gaze detection and sex categorization. *Social Neuroscience*, *3*, 69–78. https://doi.org/10.1080/17470910701563483

Correll, J., Urland, G. R., & Ito, T. A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*(1), 120–128. https://doi.org/10.1016/j.jesp.2005.02.006

Crites, S. L., Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1995). Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology*, *68*(6), 997.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Chris Gatenby, J., Gore, J. C., Banaji, M. R., Gatenby, J. C., WA, C., MK, J., CL, R., J, C. G., JC, G., & MR, B. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*(12), 806–813. https://doi.org/10.1111/j.0956-7976.2004.00760.x

Dang, T. P., Mattan, B. D., Barth, D. M., Handley, G., Cloutier, J., & Kubota, J. T. (2022). Perceiving social injustice during arrests of Black and White civilians by White police officers: An fMRI investigation. *NeuroImage*, *255*, 119153. https://doi.org/10.1016/J.NEUROIMAGE.2022.119153

Dang, T. P., Mattan, B. D., Kubota, J. T., & Cloutier, J. (2019). The ventromedial prefrontal cortex is particularly responsive to social evaluations requiring the use of person-knowledge. *Scientific Reports*, *9*(1), 5054. https://doi.org/10.1038/s41598-019-41544-z

De Dreu, C. K. W. (2010). Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Processes & Intergroup Relations*, *13*(6), 701–713.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, *25*(11), 4596–4609.

Delgado, M. R., Frank, R. H., & Phelps, E. a. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), 1611–1618. https://doi.org/10.1038/nn1575

Delplanque, S., Silvert, L., Hot, P., & Sequeira, H. (2005). Event-related P3a and P3b in response to unpredictable emotional stimuli. *Biological Psychology*, *68*(2), 107–120.

Denny, B. T., Jungles, M. L., Goodson, P. N., Dicker, E. E., Chavez, J., Jones, J. S., & Lopez, R. B. (2023). Unpacking reappraisal: a systematic review of fMRI studies of distancing and reinterpretation. *Social Cognitive and Affective Neuroscience*, *18*(1), nsad050.

Derous, E., Ryan, A. M., & Serlie, A. W. (2015). Double Jeopardy Upon Resumé Screening: When Achmed Is Less Employable Than Aïsha. *Personnel Psychology*, *68*(3), 659–696. https://doi.org/https://doi.org/10.1111/peps.12078

Di Stasi, L. L., Catena, A., Canas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968–975.

Duerden, E. G., Arsalidou, M., Lee, M., & Taylor, M. J. (2013). Lateralization of affective processing in the insula. *Neuroimage*, *78*, 159–175.

Duval, E. R., Moser, J. S., Huppert, J. D., & Simons, R. F. (2013). What's in a Face? *Journal of Psychophysiology*.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790.

Fareri, D. S., & Delgado, M. R. (2014). Differential reward responses during competition against in- and out-of-network others. *Social Cognitive and Affective Neuroscience*, *9*(4), 412–420. https://doi.org/10.1093/scan/nst006

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. https://doi.org/10.1037/0022-3514.69.6.1013

Firat, R. B., Hitlin, S., Magnotta, V., & Tranel, D. (2017). Putting race in context: Social class modulates processing of race in the ventromedial prefrontal cortex and amygdala. *Social Cognitive and Affective Neuroscience*, *12*, 1314–1324. https://doi.org/10.1093/scan/nsx052

Fitzpatrick, M. (2015). was a huge year for interracial relationships on TV. *Fusion. Retrieved from Fusion. Net*.

Forbes, C. E., Cox, C. L., Schmader, T., & Ryan, L. (2012). Negative stereotype activation alters interaction between neural correlates of arousal, inhibition and cognitive control. *Social Cognitive and Affective Neuroscience*, *7*(7), 771–781. http://scan.oxfordjournals.org/content/7/7/771.short

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology*, *46*(3), 521–530.

Fourie, M. M., Stein, D. J., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2017). Empathy and moral emotions in post-apartheid South Africa: an fMRI investigation. *Social Cognitive and Affective Neuroscience*, *12*(6), 881–892. https://doi.org/10.1093/scan/nsx019

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*, *50*(14), 3600–3611. https://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2012.09.002

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greve, D. N. (2002). *Optseq2*.

Gutsell, J. N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, *46*(5), 841–845.

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233–1235.

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, 103948.

Hackel, L. M., Zaki, J., & Van Bavel, J. J. (2017). Social identity shapes social valuation: evidence from prosocial behavior and vicarious reward. *Social Cognitive and Affective Neuroscience*, *12*(8), 1219–1228.

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion*, *8*(2), 250.

Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science*, *19*(4), 405–411.

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making*, *25*(2), 188–195.

Handley, G., Kubota, J., & Cloutier, J. (2022). Interracial contact differentially shapes brain networks involved in social and non-social judgments from faces: a combination of univariate and multivariate approaches. *Social Cognitive and Affective Neuroscience*, *17*(2), 218–230.

Handley, G., Kubota, J. T., & Cloutier, J. (2023). Reading the mind in the eyes of Black and White people: Interracial contact and perceived race affects brain activity when inferring mental states. *NeuroImage*, *269*, 119910.

Harris, L. T., Cikara, M., & Fiske, S. T. (2010). Envy as Predicted by the Stereotype Content Model: A Volatile Ambivalence. In *Envy: Theory and Research*. https://doi.org/10.1093/acprof:oso/9780195327953.003.0008

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup versus ingroup face stimuli. *Neuroreport*, *11*(11), 2351–2355. https://doi.org/10.1097/00001756-200008030-00004

Hartley, C. A., & Phelps, E. A. (2010). Changing fear: The neurocircuitry of emotion regulation. *Neuropsychopharmacology*, *35*(1), 136–146. https://doi.org/10.1038/npp.2009.121

Heekeren, H. R., Wartenburger, I., Marschner, A., Mell, T., Villringer, A., & Reischies, F. M. (2007). Role of ventral striatum in reward-based decision making. *Neuroreport*, *18*(10), 951–955.

Heining, M., Young, A. W., Ioannou, G., Andrew, C. M., Brammer, M. J., Gray, J. A., & Phillips, M. L. (2003). Disgusting smells activate human anterior insula and ventral striatum. *Annals of the New York Academy of Sciences*, *1000*(1), 380–384.

Hoffman, E. a, & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. https://doi.org/10.1038/71152

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145–E9152.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, *13*(12), 524–531. https://doi.org/10.1016/j.tics.2009.10.002

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*(6), 660–676. https://doi.org/10.1006/jesp.2000.1430

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187.

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

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective & Behavioral Neuroscience*, *5*(1), 21–36. https://doi.org/10.3758/cabn.5.1.21

Kim, S., & Lee, D. (2011). Prefrontal cortex and impulsive decision making. *Biological Psychiatry*, *69*(12), 1140–1146.

Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, *19*(10), 1280–1285.

Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., Young, L., & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage*, *161*, 9–18.

Krill, A., & Platek, S. M. (2009). In-group and out-group membership mediates anterior cingulate activation to social exclusion. *Frontiers in Evolutionary Neuroscience*, *1*(April), 1. https://doi.org/10.3389/neuro.18.001.2009

Kringelbach, M. L., & Berridge, K. C. (2009). Towards a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences*, *13*(11), 479–487. https://doi.org/10.1016/J.TICS.2009.08.006

Krosch, A. R., Tyler, T. R., & Amodio, D. M. (2017). Race and recession: Effects of economic scarcity on racial discrimination. In *Journal of Personality and Social Psychology* (Vol. 113, Issue 6, pp. 892–909). American Psychological Association. https://doi.org/10.1037/pspi0000112

Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature Neuroscience*, *15*(7), 940–948. https://doi.org/10.1038/nn.3136

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*(5), 738–752. https://doi.org/10.1016/j.jesp.2006.10.023

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795.

Kwee, I. L., & Nakada, T. (2003). Dorsolateral prefrontal lobe activation declines significantly with age Functional NIRS study. *Journal of Neurology*, *250*, 525–529.

Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *ELife*, *9*. https://doi.org/10.7554/eLife.53162

Leach, C. W., Peng, T. R., & Volckens, J. (2000). Is racism dead? Comparing (expressive) means and (structural equation) models. *British Journal of Social Psychology*, *39*(3), 449–465.

Lelkes, Y., & Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics*, *79*(2), 485–501.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences*, *1224*(1), 40–62.

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.*, *58*, 259–289.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, *8*(6), 720–722. https://doi.org/10.1038/nn1465

Ly, M., Haynes, M. R., Barter, J. W., Weinberger, D. R., & Zink, C. F. (2011). Subjective socioeconomic status predicts human ventral striatal responses to social status information. *Current Biology*, *21*(9), 794–797. https://doi.org/10.1016/j.cub.2011.03.050

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.

Mahler, S. V, Smith, K. S., & Berridge, K. C. (2007). Endocannabinoid hedonic hotspot for sensory pleasure: anandamide in nucleus accumbens shell enhances 'liking'of a sweet reward. *Neuropsychopharmacology*, *32*(11), 2267–2278.

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience*, *10*(2), 141–152.

Martin, A. K., Kessler, K., Cooke, S., Huang, J., & Meinzer, M. (2020). The right temporoparietal junction is causally associated with embodied perspective-taking. *Journal of Neuroscience*, *40*(15), 3089–3095.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1).

Mathur, V. a, Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, *51*(4), 1468–1475. https://doi.org/10.1016/j.neuroimage.2010.03.025

Mattan, B. D., Kubota, J. T., & Cloutier, J. (2017). How social status shapes person perception and evaluation: A social neuroscience perspective. *Perspectives on Psychological Science*, *12*(3), 468–507. https://doi.org/10.1177/1745691616677828

Mattan, B. D., Kubota, J. T., Dang, T. P., & Cloutier, J. (2018). External motivation to avoid prejudice alters neural responses to targets varying in race and status. *Social Cognitive and Affective Neuroscience*, *13*(1), 22–31. https://doi.org/10.1093/scan/nsx128

Mattan, B. D., Kubota, J. T., Li, T., Dang, T. P., & Cloutier, J. (2018). Motivation modulates brain networks in response to faces varying in race and status: A multivariate approach 3. *ENeuro*, *5*(4). https://doi.org/e0039-18.2018

Mattan, B. D., Wei, K. Y., Cloutier, J., & Kubota, J. T. (2018). The social neuroscience of race-based and status-based prejudice. *Current Opinion in Psychology*, *24*, 27–34. https://doi.org/10.1016/j.copsyc.2018.04.010

McCarthy, J. (2021). US approval of interracial marriage at new high of 94%. *Gallup. Com*.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*(3), 285–299.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*, 655–667. https://doi.org/10.1007/s00429-010-0262-0

Merritt, C. C., MacCormack, J. K., Stein, A. G., Lindquist, K. A., & Muscatell, K. A. (2021). The neural underpinnings of intergroup social cognition: an fMRI meta-analysis. *Social Cognitive and Affective Neuroscience*, *16*(9), 903–914. https://doi.org/10.1093/scan/nsab034

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315.

Molapour, T., Golkar, A., Navarrete, C. D., Haaker, J., & Olsson, A. (2015). Neural correlates of biased social fear learning and interaction in an intergroup context. *NeuroImage*, *121*, 171–183.

Molenberghs, P., Ogilvie, C., Louis, W. R., Decety, J., Bagnall, J., & Bain, P. G. (2015). The neural correlates of justified and unjustified killing: an fMRI study. *Social Cognitive and Affective Neuroscience*, *10*(10), 1397–1404.

Moll, J., R, de O.-S., FT, M., FA, I., & IE, B. (2005). The moral affiliations of disgust: a functional MRI study. *Cogn. Behav. Neurol.*, *18*, 68.

Morrison, S., Decety, J., & Molenberghs, P. (2012). The neuroscience of group membership. *Neuropsychologia*, *50*(8), 2114–2120.

Mullins, D. (2014). Advertising features more gay, interracial families. *Aljazeera America*.

Ochsner, K. N., Silvers, J. A., & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, *1251*(1), E1–E24.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*(5735), 785–787.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience*, *2*(1), 3–11. https://doi.org/10.1093/scan/nsm005

Ong, K. (2020). Ideological convergence in the extreme right. *Counter Terrorist Trends and Analyses*, *12*(5), 1–7.

Pandey, E. (2022). *The rise of interracial marriage—and its approval rating*. Axios. Retrieved from https://www. axios. com/2022/09/07/approval-of ….

Pape, R. A. (2021). *Understanding American Domestic Terrorism Mobilization Potential and Risk Factors of a New Threat Trajectory*. https://d3qi0qp55mx5f5.cloudfront.net/cpost/i/docs/americas_insurrectionists_online_2021_04_06.pdf

Parker, K., & Barroso, A. (2021). *In Vice President Kamala Harris, we can see how America has changed*.

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love." *Social Justice Research*, *26*(1), 81–96.

Payne, B. K. (2001). Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181.

Peciña, S., & Berridge, K. C. (2005). Hedonic hot spot in nucleus accumbens shell: where do μ-opioids cause increased hedonic impact of sweetness? *Journal of Neuroscience*, *25*(50), 11777–11786.

Peciña, S., Smith, K. S., & Berridge, K. C. (2006). Hedonic hot spots in the brain. *The Neuroscientist*, *12*(6), 500–511.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *24*(57), 27–53. https://doi.org/10.1146/annurev.psych.56.091103.070234

Phelps, E. A., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia*, *41*(2), 203–208. https://doi.org/10.1016/S0028-3932(02)00150-1

Phelps, E. A., & LeDoux, J. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*(5), 729–738. https://doi.org/10.1162/089892900562552

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.

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & :. J. N. Shelton. (2003). An fMRI examination of the impact of interracial contact on executive function. *Nature Neuroscience*, *6*, 1323–1328. https://doi.org/10.1038/nn1156

Richeson, J. A., Todd, A. R., Trawalter, S., & Baird, A. A. (2008). Eye-gaze direction modulates race-related amygdala activity. *Group Processes & Intergroup Relations*, *11*(2), 233–246. https://doi.org/10.1177/1368430207088040

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447.

Roberts, K. L., & Hall, D. A. (2008). Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks. *Journal of Cognitive Neuroscience*, *20*(6), 1063–1078.

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in Neurobiology*, *86*(3), 216–244.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*, 147–156. https://doi.org/10.1016/j.tics.2012.01.005

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., Schmidt-Wilcke, T., & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory-of-mind: A neuroimaging study using cartoon picture stories. *Human Brain Mapping*, *36*(1), 29–39.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34.

Sharp, D. J., Scott, S. K., & Wise, R. J. S. (2004). Monitoring and the controlled processing of meaning: distinct prefrontal systems. *Cerebral Cortex*, *14*(1), 1–10.

Shkurko, A. V. (2013). Is social categorization based on relational ingroup/outgroup opposition? A meta-analysis. *Social Cognitive and Affective Neuroscience*, *8*(8), 870–877.

Simi, P., Blee, K., DeMichele, M., & Windisch, S. (2017). Addicted to hate: Identity residual among former white supremacists. *American Sociological Review*, *82*(6), 1167–1187.

Simi, P., & Futrell, R. (2010). *American Swastika: Inside the white power movement's hidden spaces of hate*. Rowman & Littlefield.

Simi, P., Futrell, R., & Bubolz, B. F. (2016). Parenting as activism: Identity alignment and activist persistence in the white power movement. *The Sociological Quarterly*, *57*(3), 491–519.

Simmons, A., Matthews, S. C., Stein, M. B., & Paulus, M. P. (2004). Anticipation of emotionally aversive visual stimuli activates right insula. *Neuroreport*, *15*(14), 2261–2265.

Singer, T., Seymour, B., JP, O., KE, S., RJ, D., & CD, F. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466.

Skinner, A. L., & Hudac, C. M. (2017). "Yuck, you disgust me!" Affective bias against interracial couples. *Journal of Experimental Social Psychology*, *68*, 68–77. https://doi.org/https://doi.org/10.1016/j.jesp.2016.05.008

Skinner, A. L., & Rae, J. R. (2019). A robust bias against interracial couples among White and Black respondents, relative to multiracial respondents. *Social Psychological and Personality Science*, *10*(6), 823–831.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default

mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510.

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

Stanley, D. A., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, *17*(2), 164–170.

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.

Stillwell, A., & Lowery, B. S. (2021). Gendered racial boundary maintenance: Social penalties for White women in interracial relationships. *Journal of Personality and Social Psychology*, *121*(3), 548.

Straube, T., & Miltner, W. H. R. (2011). Attention to aversive emotion and specific activation of the right insula and right somatosensory cortex. *Neuroimage*, *54*(3), 2534–2538.

Tajfel, H. (1982). *Social identity and intergroup behavior*. Cambridge University Press.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, *56*, 65.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science*, *323*(5916), 937–939.

Tamir, D. I., & Hughes, B. L. (2018). Social rewards: from basic social building blocks to complex social behavior. *Perspectives on Psychological Science*, *13*(6), 700–717.

Telzer, E. H., Ichien, N., & Qu, Y. (2015). The ties that bind: Group membership shapes the neural correlates of in-group favoritism. *NeuroImage*, *115*, 42–51.

Tupak, S. V, Dresler, T., Guhn, A., Ehlis, A.-C., Fallgatter, A. J., Pauli, P., & Herrmann, M. J. (2014). Implicit emotion regulation in the presence of threat: neural and autonomic correlates. *Neuroimage*, *85*, 372–379.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*, 55–61. https://doi.org/10.1038/nrn3857

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

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*(6), 1037–1050.

Wagner, D. D., Altman, M., Boswell, R. G., Kelley, W. M., & Heatherton, T. F. (2013). Self-regulatory depletion enhances neural responses to rewards and impairs top-down control. *Psychological Science*, *24*(11), 2262–2271.

Wang, Y., Schubert, T. W., & Quadflieg, S. (2019). Behavioral and neural evidence for an evaluative bias against other people's mundane interracial encounters. *Social Cognitive and Affective Neuroscience*, *14*(12), 1329–1339.

Weisel, O., & Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, *60*, 110–120.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56–63. https://doi.org/10.1111/j.0956-7976.2005.00780.x

Wicker, B., Keysers, C., Plailly, J., Royet, J., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*, 671–684. https://doi.org/10.3758/brm.42.3.671

Willis, M. L., Palermo, R., Burke, D., McGrillen, K., & Miller, L. (2010). Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. *Neuropsychologia*, *48*(7), 2182–2187.

Windisch, S., Scott Ligon, G., & Simi, P. (2019). Organizational [dis] trust: Comparing disengagement among former left-wing and right-wing violent extremists. *Studies in Conflict & Terrorism*, *42*(6), 559–580.

Windisch, S., Simi, P., Blee, K., & DeMichele, M. (2020). Measuring the Extent and Nature of Adverse Childhood Experiences (ACE) among Former White Supremacists. *Terrorism and Political Violence*, 1–22. https://doi.org/10.1080/09546553.2020.1767604

Ying, X., Luo, J., Chiu, C., Wu, Y., Xu, Y., & Fan, J. (2018). Functional dissociation of the posterior and anterior insula in moral disgust. *Frontiers in Psychology*, *9*, 860.

Young, L., & Saxe, R. (2009). An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405. https://doi.org/10.1162/jocn.2009.21137

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience*, *6*(3), 302–315.

Zhang, M., & Peng, Y. (2023). Anterior insula and dorsal anterior cingulate cortex as a hub of self-regulation: combining activation likelihood estimation meta-analysis and meta-analytic connectivity modeling analysis. *Brain Structure and Function*, *228*(5), 1329–1345.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron*, *58*(2), 273–283. https://doi.org/10.1016/j.neuron.2008.01.025

Appendix

IRB/HUMAN SUBJECT APPROVAL

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

DATE:	February 27, 2025
TO:	Jennifer Kubota, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[1778845-13] Affective Impression Formation
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
EFFECTIVE DATE:	February 27, 2025
NEXT REPORT DUE:	August 24, 2025
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(4,7)*

Thank you for your Amendment/Modification 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 August 24, 2025. 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. Please include the study title and reference number in all correspondence with this office.

INSTITUTIONAL REVIEW BOARD

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

DATE:	July 1, 2024
TO:	Jennifer Kubota, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[1444806-9] Affective Stimuli Pilot
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
EFFECTIVE DATE:	July 1, 2024
NEXT REPORT DUE:	July 29, 2025
REVIEW TYPE:	Administrative Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your Continuing Review/Progress Report 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 Administrative 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 July 29, 2025. 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. Please include the study title and reference number in all correspondence with this office.

INSTITUTIONAL REVIEW BOARD

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

DATE:	August 6, 2024
TO:	Jennifer Kubota, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[1467259-8] Social Stimuli Pilot
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
EFFECTIVE DATE:	August 6, 2024
NEXT REPORT DUE:	September 3, 2025
REVIEW TYPE:	Administrative Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your Continuing Review/Progress Report 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 Administrative 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 September 3, 2025. 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. Please include the study title and reference number in all correspondence with this office.

INSTITUTIONAL REVIEW BOARD