THE INFLUENCE OF RACE ON BEHAVIOR AND THE NEURAL CORRELATES OF
EXPECTANCY DURING AN ECONOMIC DECISION GAME

A Dissertation

presented to

the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

MEREDITH P. LEVSEN

Dr. Bruce Bartholow, Dissertation Supervisor

MAY 2018

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE INFLUENCE OF RACE ON BEHAVIOR AND THE NEURAL CORRELATES OF
EXPECTANCY DURING AN ECONOMIC DECISION GAME

presented by Meredith P. Levsen,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Bruce Bartholow

Professor Laura Scherer

Professor Clinton Davis-Stober

Professor Jason Craggs

To my parents, Dr. Stewart Johnson and Jane Kestenbaum, who inspire me and instilled in me the skills I needed to be successful.

To my sister, Emma Johnson – remember, you are so pretty.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Bruce Bartholow, for his guidance and support throughout all stages of the development and execution of the work presented here. I would also like to thank the members of my dissertation committee, Dr. Laura Scherer, Dr. Clint Stober-Davis, and Dr. Jason Craggs, for providing helpful comments and suggestions to improve the quality of this work. I would further like to thank the undergraduate research assistants who helped with data collection. I am also indebted to the participants, without whom, this project could not have come to fruition. I owe a very special thanks to the members of the Bartholow Lab for enduring countless presentations of the data, and for providing me with methodological guidance, intellectual input, emotional support, and friendship along the way.

I am also very appreciative of Matt and Janis Levsen, who have so kindly welcomed me into their family. Also, thank you to Kris Wallace and everyone at Columbia Equestrian Center for providing me with such a wonderful place to go when I needed a reprieve from the stress of graduate school. Additionally, I would like to thank Gatsby and Dallas for keeping me company during the late nights I spent working on this manuscript – it would have been far more unpleasant without them. Finally, I owe a huge thanks to Tyler Levsen for enthusiastically listening to practice presentations of this data, providing me with sustenance and keeping our home in order when my academics took priority, and for being a bottomless well of emotional support.

Table of Contents

Acknowledgements	ii
List of Figures.....	v
List of Tables	vi
Abstract.....	vii
Introduction.....	1
Social Expectancy Violations and Behavioral Economics Games	3
Differential Treatment Based on Race during Social Economic Games	4
Neurocognitive Mechanisms of Expectancy and Social Categorization	7
The Current Study	14
Method	18
Participants	18
Exclusion Criteria Considerations.....	18
General Procedure	19
Trust Game Paradigm.....	20
Stimuli Database Project Deception.....	23
Questionnaires	25
Funneled Debriefing.....	26
Data Reduction and Cleaning.....	26
Results	33
Analyses of Primary Hypotheses	33
Exploratory Analyses of Secondary Hypotheses	38
Discussion	43
Behavior during the Trust Game.....	43
ERP Correlates of Expectancy (RewP) and Motivational Salience (P3).....	44
Exploratory Examination of the ERP Correlate of Social Categorization (P2)	47
Limitations and Future Directions.....	49
Conclusion.....	50
References.....	52
Figures.....	64
Appendix A.....	77
Appendix B.....	79
Appendix C	82
Appendix D.....	84

Supplemental Methods	88
Vita	92

List of Figures

FIGURE 1.	64
FIGURE 2.	66
FIGURE 3.	67
FIGURE 4.	69
FIGURE 5.	71
FIGURE 6.	72
FIGURE 7.	73
FIGURE 8.	75

List of Tables

Supplemental Table S1	88
-----------------------------	----

THE INFLUENCE OF RACE ON BEHAVIOR AND THE NEURAL CORRELATES
OF EXPECTANCY DURING AN ECONOMIC DECISION GAME

Meredith P. Levensen

Dr. Bruce D. Bartholow, Dissertation Supervisor

Abstract

I investigated how another person's race influences resource sharing during a social economic decision-making game that depends on trustworthiness. Seventy-eight university students played a behavioral economics Trust Game with players of different races while EEG was recorded. I predicted that negative stereotype-based expectancies of Black players would cause participants to be more trusting of White than Black players, and also that these expectancies would be evident in the event-related potential (ERP) component, RewP, which is sensitive to expectancy violations (or prediction errors). Results largely conformed to predictions, in that White participants shared more money with White players than with Black players; this effect was weaker (and not statistically significant) when data from the full sample, including 11 non-White participants, was used in the model. This preferential treatment of White players was more pronounced in individuals who held more biased racial attitudes. Additionally, RewP was greater for feedback from Black relative to White players on trials when offer amounts were higher than average, implying that when feedback was most salient (because more money was at stake), it was more unexpected when Black players reciprocated generosity, and less expected when they did not. Exploratory analyses of an ERP component associated with racial categorization of faces, the P2, was larger to Black versus White faces, as in previous studies. Of interest here, the race effect in P2 amplitude was moderated by racial

attitudes, such that less-biased individuals showed larger race differences in the P2, suggesting that these participants allocated more attention to social categories in an attempt to respond in a less biased way. These novel results were not entirely consistent with my original hypotheses, and thus more research is necessary to replicate and help explain the present results.

Introduction

Although racial inequality has been declining in recent decades (Fischer & Hout, 2006), the disadvantages of being an African American in the U.S. are enormously impactful and pervasive, affecting employment, health, income, wealth, education, socioeconomic mobility, and education, to name a few (Pager & Shepherd, 2008; Stanford Center for the Study of Poverty and Inequality, 2017; Williams & Mohammed, 2013). For example, as of 2015 one in four Blacks are below the poverty line, whereas one in ten Whites are in poverty (Burton, Mattingly, Pedroza, & Welsh, 2017). It is important to note that the present state of racial inequality is not merely the vestiges of America's racist past, filled with open practices of discrimination. There exists evidence that discrimination is still widespread. For instance, one experiment found that White males were twice as likely to receive entry-level job offers as equally qualified Black males (Pager, Western, & Bonikowski, 2009). It is believed that racial bias – be it conscious thoughts, or subtle feelings beyond the scope of awareness – drives individual behavior and decision-making in such a way that gives rise to the present state of racial inequality (e.g., Payne & Cameron, 2010).

One way that beliefs and attitudes translate into actions and behavior is through *expectancies*. Expectancies are automatic assumptions that people make based on previous knowledge or experiences. These assumptions fill in gaps in our knowledge about novel stimuli by allowing us to apply previously learned knowledge about similar stimuli (Roese & Sherman, 2007). This increases mental efficiency by allowing us to not fully process everything we encounter (Roese & Sherman, 2007). Expectancies can be derived from and triggered by many different aspects of both our internal and external

environment (Roese & Sherman, 2007). Activation of expectancies can happen automatically, seemingly without thought or exerted effort, which is precisely why they are so efficient and useful. For example, assume a person was entering a post office building she had never seen before. It would most likely not require much effort for her to find the entrance because she would know to look for a door, based on her previous experience with entering public buildings. Expectancies influence our perceptual processes as they help us seamlessly navigate the world around us, simultaneously shaping and being shaped by our emotions, judgments, and attitudes.

When applied to social psychology, social expectancies can be considered prior expectations or predictions about other people's behavior. Often, we transform perceptual information into a categorical identity, which then allows us to hypothesize what other people might do based on their category membership (e.g., Jones, 1990). This can actually be quite helpful – being able to make quick judgments about people is important and necessary for survival. For example, if we encounter an angry-looking person lurking in a dark alley with a large butcher knife in hand, it is imperative that we be able to make assumptions about this person based on their threatening appearance to help us fill in the gaps in our knowledge. Doing so helps us anticipate that this person is potentially dangerous so that we can act quickly to remove ourselves from harm's way.

Although predominantly helpful, expectancies are sometimes harmful and can interfere with goal pursuit. Sometimes, expectancies can reflect attitudes or beliefs that the holder would never explicitly endorse. For example, social group stereotypes can be used to inform expectancies about individual group members, and can ultimately shape behavior and interactions with outgroup members (e.g., Jussim, Coleman, & Lerch,

1987). A topical example of how stereotype-based expectancies shape behavior without intention is the many instances of police shooting unarmed Black men. In many of these cases, it seems that the police made erroneous assumptions about the level of threat posed by these men. Advancing our understanding of how stereotype-based expectancies infiltrate thoughts and actions may eventually aid in preventing such deadly consequences.

Social Expectancy Violations and Behavioral Economics Games

In spite of the ubiquitous influence of expectancies on our actions and cognitions, in many cases, we remain unaware of their influence unless they fail – that is, when we encounter information that violates our expectancies (Roese & Sherman, 2007). In such cases, we may be required to engage in cognitive processes to update our expectancies in an attempt to reconcile our violated expectancy with newly introduced, conflicting information. Relative to expectancy confirmations, *expectancy violations* are associated with increased affective arousal (e.g., Jussim et al., 1987) and cognitive processing (Srull & Wyer, 1989; Stangor & McMillan, 1992). These consequences aid in facilitating learning through experiences (i.e., reinforcement learning), signaling that templates stored in long-term memory need to be updated. In behavioral economics, expectancy violations are known as *prediction errors*, defined as the difference between the expected value of the outcome and its actual value.

Recent work has begun to use social behavioral economics games to investigate the consequences of social expectancies and expectancy violations. During a typical behavioral economics gambling task, participants are often asked to choose between different options with varying probabilities of receiving a specified reward or punishment

(e.g., Option A = 25% chance of winning \$10 vs. Option B = 40% chance of winning \$5). In some tasks, the participant may choose to risk losing some money in order to earn an even greater reward. Instead of the risk associated with different options being explicitly specified by the experimenter, social behavioral economics games require participants to infer risk based on social information (e.g., a picture of the other player) provided by the experimenter. This inferred risk then can be measured via the behavioral output from the participant during the task (i.e., the choices they select). Granted, there are factors that can contribute to a participant's response other than inferred risk based on attributes of the other player (e.g., risk aversion). Nonetheless, social behavioral economics games provide social psychologists with an interesting, uniquely behavior-oriented approach to quantifying and measuring tangible consequences of social expectancies in the midst of a social interaction.

Differential Treatment Based on Race during Social Economic Games

Recently, social economic decision-making games have been utilized to characterize behavioral influences of expectancies elicited by one's racial group. (Kubota, Li, Bar-David, Banaji, & Phelps, 2013) designed a modified single-shot (meaning other players are only encountered once) version of the ultimatum game (UG; Guth, Schmittberger, & Schwarze, 1982), a two-player game that emphasizes judgments of fairness. In this version of the UG, the participants played as the "responder." The other players (represented by pictures of different faces that varied by race) played the role of the "proposer". At the beginning of each trial, a sum of money is allotted to the proposer to divide between him- or herself and the other player. Offers made by the proposer can vary in fairness. Once an offer is made, the responder can either "accept" or

“reject” the offer. If accepted, the money is distributed according to the proposal; if rejected, both players receive nothing. Participants were told that the other players (i.e., the proposers) were real people whose offers had been previously recorded. They found that participants accepted more offers from White proposers than Black proposers (cf. Mendoza, Lane, & Amodio, 2014). Furthermore, they found that this effect was moderated by individual differences in implicit racial bias, such that those with greater anti-Black bias were more likely to reject offers from Black than White proposers.

Stanley, Sokol-Hessner, Banaji, and Phelps (2011) sought to test whether implicit racial bias predicted the discrepancy in behavior toward Black and White counterparts during a modified version of a social behavioral economics game known as the Trust Game (TG; Berg, Dickhaut, & McCabe, 1995). During the two-player version of this game, one player is the “investor” and the other the “trustee.” Similar to the UG, each round begins with the investor receiving a sum of money, and then he or she must decide how to divide the money between him/herself and the trustee. Whatever they do not share with the trustee, they keep, and the amount they shared immediately multiplies by a constant value (e.g., 4) that varies with different versions of the TG. Then the trustee gets to decide between two options: keep the total amount for themselves, or reciprocate the investor’s generosity by returning half of the multiplied sum to the investor.

In the experiment, participants played the role of investor with a new partner during each trial. During each trial, participants were presented with a picture of an emotionally neutral face and asked how much they would like to share with that person. After responding, participants were not given feedback regarding whether their partner had decided to reciprocate or not. The authors were concerned that receiving feedback

might affect subsequent decisions. They found no differences in the amount participants shared with White and Black trustees.

However, they did find that this effect was moderated by individual differences in implicit bias. Implicit bias is generally defined as a preference or attitude that arises automatically during earlier stages of perception before more reflective, higher levels of processing have had time to take place (e.g., Stanley, Phelps, & Banaji, 2008). These attitudes are influenced and formulated by past experiences, and they can sometimes be introspectively inaccessible (e.g., Greenwald & Banaji, 1995). Implicit bias is typically measured indirectly using fast-paced reaction time social cognition tasks that rely on automatic categorical associations. The authors measured implicit bias using the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) to assess individual differences in the tendency to associate pleasant and unpleasant words with pictures of Black and White faces. They found that more biased participants made higher offers to White than Black partners, suggesting that bias is associated with giving preferential treatment and trust to White players. These results indicate that stereotype-based expectancies interact with individual differences in implicit racial bias to inform economic decisions and related evaluations of trustworthiness.

In a follow-up study, Stanley et al. (2012) sought to test the neural correlates of the social judgment decision-making process using an almost identical paradigm. This time, instead of using faces of people from a wide age range, the face stimuli representing the Trustees were of college age. However, the behavioral effects Stanley et al. (2011) found previously did not replicate. It is possible that playing against college-age peers from a nearby university removed a degree of social distance that was present during the

first study. This latter study might highlight a boundary condition on the race effect observed during economic decisions of trustworthiness (cf. Kubota et al., 2013).

Although progress has been made to advance our understanding of why and when racial bias influences decision-making within the context of financial transactions, the few studies that have been conducted to date have only begun to scratch the surface. Further research is necessary in order to generate a clearer picture of how and when expectancies drawing from racial stereotypes affect social interactive economic decisions.

From the perspective of a researcher interested in understanding racial bias in economic decisions, there are many aspects of social economic decision-making games that are alluring; and they offer a promising, novel way of exploring how race interacts with decision-making related to resources. This approach has especially important implications in terms of potentially providing insight into racial disparity that exists in the United States. It is easy to imagine how similar expectancy processes that lead an employer to overlook applications from Black prospective employees might also lead someone to entrust less money to a Black player during the TG. So far, these methods have been underutilized for studying racial bias.

Neurocognitive Mechanisms of Expectancy and Social Categorization

Behavior during social economic decision-making games can help us to further understand and probe racial bias. However, like all experimental tasks, relying solely on behavioral responses during these games has its limitations. For example, in the TG, how much an individual shares with their counterpart is not a pure representation of how much they trust the other player with money. It is possible that other factors affect the investor's decisions during the task. For example, perhaps their "true response" would be

to share very little with a Black trustee because they do not trust that they would reciprocate, but feelings of “White guilt” compel them instead to share the entire sum of money with the trustee. In this scenario, the amount of money shared would actually be a poor estimate of trust. This could be addressed by asking participants directly how they made their decisions. However, responses on such measures may also be limited or inaccurate. People may not be completely forthcoming when discussing personal beliefs about a sensitive topic such as racial bias. Additionally, people are sometimes inaccurate in their introspective self-reports of their own cognitive processes (Kunda, 1990; Nisbett & Wilson, 1977; Zajonc, 1980).

Applying psychophysiological measures is one solution that would allow examination of underlying mental processes as they unfold throughout the task. Measuring an individual’s neural responses can offer a second source of information that can be used to corroborate the story told by the behavioral measures. Returning to the above example, examining the neural correlates of expectancies could provide insight into the beliefs the person held about a given counterpart during the TG (independent of the actual behavioral decision they made). Fortunately, there are event-related potential (ERP) components of the electroencephalography (EEG) signal that have been associated with expectancy during reward learning tasks. EEG technology provided a cost-effective way of examining the neural underpinnings of expectancy.

ERPs associated with expectancy. Previous research has shown that two components of the ERP, the Reward Positivity (RewP) and the P3, are particularly sensitive to expectancies. The RewP (Proudfit, 2015) is a positive deflection in electrophysiological signal, maximal at frontocentral electrode sites, which emerges 200-

400 ms following performance feedback. The RewP is thought to reflect activity from regions in the basal ganglia (Becker, Nitsch, Miltner, & Straube, 2014; Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011; Foti, Carlson, Sauder, & Proudfit, 2014; Foti, Weinberg, Dien, & Hajcak, 2011), an area involved in reward processing. RewP is hypothesized to reflect a prediction error (or expectancy violation) - that is, the extent to which an outcome or feedback diverges from one's expectations (e.g., Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Nieuwenhuis, Holroyd, Mol, & Coles, 2004). The RewP signal seems to be mostly driven by positive reward feedback specifically, exhibiting larger effects within these manipulations (e.g., Kreussel et al., 2011; for review, see Proudfit, 2015). When the feedback outcome is worse than expected, this generates a less positive RewP signal; however, when the outcome is better than expected, the RewP signal is more positive. Accordingly, RewP has been demonstrated to respond to factors such as valence (reward, punishment), magnitude, and probability during reward learning tasks (see Kreussel et al., 2012). These attributes make RewP an ideal candidate for assessing expectancy violations and inferring expectancy during social economic games.

A growing body of evidence suggests that RewP is sensitive to various social cues (e.g., attractiveness) that can be used to generate predictions about others' behavior during tasks like the Trust Game (Campanhã, Minati, Fregni, & Boggio, 2011; Chen et al., 2012; Osinsky, Mussel, Öhrlein, & Hewig, 2014; Qu, Wang, & Huang, 2013; Tortosa, Lupiáñez, & Ruz, 2013; Yin Wu, Leliveld, & Zhou, 2011; Yin Wu, Zhou, van Dijk, Leliveld, & Zhou, 2011). Previous research has found that during the UG, RewP is more positive for fair than unfair offers (e.g., Boksem & De Cremer, 2010). Furthermore,

the RewP signal is associated with the negative affect associated with unfair offers, and has been demonstrated to predict the subsequent rejection of offers (Hewig et al., 2011). Additionally, accumulating evidence indicates that the RewP amplitude is also sensitive to social expectancies within the context of social economic decision-making. Osinsky and colleagues (2014) found that, in the role of responder, the RewP component was affected by both the magnitude of the offer and the learned reputation of particular proposers based on previous offers. Interestingly, they found that as the reputations of certain proposers became more established (i.e., learned) as the experiment progressed, the RewP latency actually shifted, so that it began to occur upon presentation of the proposer's face.

Far fewer studies have examined RewP during the TG. One study conducted by Chen and colleagues (2012) tested the effect of facial attractiveness on expectancy violations to feedback, as reflected in RewP. The authors found that playing against more attractive partners yielded larger differential RewP signals for positive and negative feedback. They reasoned that because people implicitly expect attractive people to be more trustworthy, feedback contradicting this preconception elicited a greater prediction error. This is a promising example of how RewP can be used as evidence of divergent *a priori* predictions held by participants about the behavior of the other players based on physical facial features. If physical features such as attractiveness can influence expectancies observed in RewP, then it is feasible that race could as well.

In summary, the extant literature suggests that RewP is sensitive to social expectancies that result in divergent prediction errors. The probability of the other player's behavior (e.g., whether they will share or keep the money as the trustee during

the TG) is generally unknown during these behavioral economics games. However, the probability of the other player's behavior can be inferred by drawing on what information is available about the other player, such as their reputation (e.g., Osinsky et al., 2014) or physical appearance (e.g., Chen et al., 2012). Once the actual behavior of the other player is revealed, differences in RewP elicited by this feedback can unveil differences in *a priori* expectations. If participants' *a priori* expectations about the other player's behavior are the same regardless of the other player's race, then there should be no difference in the RewP amplitude when playing against players of different races. However, if the other player's race factors into participants' expectations of the other players' behavior, RewP ought to differ by race. Examining RewP was of primary interest when this experiment was designed.

The P3 (a.k.a., P300, or P3b; see Polich, 2007) is a parietally maximal, positive-going ERP waveform that peaks around 300 to 800 ms after stimulus onset (for review, see Nieuwenhuis, Aston-Jones, & Cohen, 2005; Polich, 2007). P3 is currently thought to reflect the extent of motivated attention elicited by a task-relevant stimulus (Nieuwenhuis et al., 2005), corresponding to its incentive value (Begleiter, Porjesz, Chou, & Aunon, 1983). In line with this theory, P3 has attributes that make it relevant to the current investigation.

P3 has been found to respond to social expectancy violations (Bartholow, Dickter, & Sestir, 2006; Bartholow, Fabiani, Gratton, & Bettencourt, 2001; Bartholow, Pearson, Gratton, & Fabiani, 2003). For example, exposure to sentences with counter-stereotypic noun-pronoun agreement (such as, "Our aerobics instructor gave *himself* a break") produces larger P3s, compared to similar sentences involving stereotype-consistent word

pairs (Osterhout, Bersick, & McLaughlin, 1997). Moreover, the same study found that P3 amplitude in response to such sentences was moderated the extent to which participants found these kinds of statements to be unusual.

In the context of reward learning and decision-making tasks (e.g., the types of gambling tasks cited above for RewP), P3 is elicited by outcome feedback, peaking soon after RewP (at a more posterior scalp location). P3 is sensitive to probability, with less frequently occurring, more novel stimuli generating larger P3 amplitudes (Polich, 2007). Robust evidence suggests that P3 fluctuates with the magnitude of rewards and losses, with a greater amplitude reflecting larger magnitude (e.g., Sato et al., 2005; Yan Wu & Zhou, 2009). There is conflicting evidence regarding whether P3 encodes feedback valence (i.e., gains vs. losses); some studies found no evidence that P3 differs by valence (Sato et al., 2005; Yeung & Sanfey, 2004), while others found P3 is greater for positive than negative feedback (Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Holroyd, & Simons, 2007; Holroyd, Hajcak, & Larsen, 2006; Yeung, Holroyd, & Cohen, 2005). The literature also contains conflicting results on the role of outcome expectancies, with some studies indicating that P3 is greater for unexpected compared to expected feedback (e.g., Hajcak et al., 2005, 2007; Pfabigan, Alexopoulos, Bauer, & Sailer, 2010) and others indicating that P3 is modulated by expectancies only some of the time (Chase, Swainson, Durham, Benham, & Cools, 2010; Yan Wu & Zhou, 2009). Consistent with the motivational salience hypothesis of P3, it has been proposed that P3 is greater when feedback received has implications for modifying future behavior (Chase et al., 2010; Yeung & Sanfey, 2004). This hypothesis is consistent with a more general theory of P3 linking its amplitude to decision-making processes that have implications for

current goals (Nieuwenhuis et al., 2005), and might account for the apparently conflicting results in this literature.

In spite of P3's apparent sensitivity to social expectancy violations, it seems that P3 (in response to feedback during a decision-making task) may only reflect expectancy when the feedback motivates the individual to adjust their behavior. As a measure of motivational salience, P3 is relevant to the current experiment. However, the relevance of motivational salience is secondary to expectancies. Additionally, this experiment was not designed to examine motivational salience specifically. Although investigation of P3 was deemed important and interesting, subsequent analyses involving P3 were considered strictly exploratory.

Social categorization of faces. Like P3, face processing ERPs were not directly relevant to my primary hypotheses about bias and expectancies. Nevertheless, given that a growing literature suggests that a particular ERP component elicited by faces, the P2, is sensitive to social categorization (Dickter & Bartholow, 2007; Ito & Urland, 2003, 2005; Kubota & Ito, 2007), it is potentially useful to examine P2 amplitude in a TG in which racial categories are expected to play an important role. As with the P3, this examination is considered exploratory.

The P2 component is a positive deflection that emerges 150-250 ms after face onset in the midline central parietal region. Broadly, P2 reflects automatic attention allocation, whereby an enhanced P2 signal to a visual stimulus (faces and non-faces alike) generally signifies greater attention directed to that stimulus (Luck, 2005). P2 in response to face stimuli is thought to arise from neural activity involved in automatically attending to social categories (Ito, 2011). Research has consistently demonstrated that

outgroup faces elicit larger P2s than ingroup faces (e.g., Amodio, 2010; Dickter & Bartholow, 2007; Dickter & Kittel, 2012). This effect is not influenced by task relevance (He, Johnson, Dovidio, & McCarthy, 2009; Ito & Urland, 2003, 2005; Kubota & Ito, 2007; but see Amodio, 2010) or the context of different tasks (Correll, Urland, & Ito, 2006; Dickter & Bartholow, 2007; Willadsen-Jensen & Ito, 2008).

Correll and colleagues (2006) investigated the relationship between the P2 race effect (i.e., how much larger P2 is for Black/outgroup than White/ingroup faces) and behavioral measures of racial bias. Participants played a first-person shooter game designed to measure implicit racial bias (Correll, Park, Judd, & Wittenbrink, 2002). During each trial, a picture of a White or Black male appears onscreen holding an object (e.g., a cell phone, soda can, etc.). The goal of the game is to “shoot” (via button press) when the male is holding a gun, and press a different button in order to “not shoot” if the person is not holding a gun. A larger race differential in P2 during the shooter task was associated with greater racial bias exhibited during the shooter task and higher personal endorsements of violent stereotypes of Black people (measured via a questionnaire). Additionally, a stronger belief in violence-based, cultural stereotypes of Black people was associated with greater racial bias during the shooter task, and this relationship was mediated by the size of the P2 race effect. This suggests that the outgroup race effect typically observed in P2 is perhaps a promising measure of racial bias that emerges within 200 ms of seeing a face.

The Current Study

The overarching goal of the current study was to investigate how another person’s race influences expectancies during economic decisions, by examining the neural and

behavioral manifestations of these race-based expectancies. The present experiment used the Trust Game (TG) because of the real world implications this game has in terms of trusting individuals with money. In line with previous TG experiments, the participants played the role of the investor while playing a series of other players (trustees) whom they believed to be real people (Stanley et al., 2012, 2011). Participants specified how much money they wanted to share with each of many players who were either White or Black. By sharing, it was possible for them to earn more money, but they also ran the risk of losing that money as well. Participants believed the outcome was determined by the decision made by other player. This experiment extends previous work (Stanley et al., 2012, 2011) by giving participants feedback about the other player's decision throughout the task.

I predicted that negative stereotypes about Black people would give rise to the expectancy that Black players are less likely to reciprocate offers. Following from this notion is the hypothesis that participants would share less money with Black players than with White players, and that this racial disparity would be more pronounced in individuals who hold more racially biased attitudes and beliefs.

Consistent with the expectancy that Black players are less likely to reciprocate, I anticipated finding evidence of RewP encoded corresponding expectancy violations (i.e., prediction errors) in response to feedback during the game. I expected to observe a main effect of race, such that RewP is more positive following feedback from Black than White players for both wins and losses. When feedback indicates that an outcome is *worse* than expected, this produces a *smaller* RewP magnitude (i.e., unexpected loss < expected loss); however, when the feedback indicates that the result is *better* than

expected, the RewP magnitude is *greater* (i.e., expected gain < unexpected gain). Upon receiving either positive or negative feedback (i.e., when partners reciprocated, or kept the money, respectively), I predicted RewP amplitude would be greater for Black than White partners. Because unexpected losses are smaller than expected losses, this would mean that it was more unexpected when White players kept the money. However, because unexpected gains are greater than expected gains, this same result would imply that the expectation of reciprocation was greater for White than for Black partners. Additionally, I predicted this effect would be exaggerated in individuals with more racially biased attitudes.

In exploratory analyses of P3, I expected to observe a Race x Feedback interaction, whereby P3 would be greater when White partners kept the money compared to Black partners. This result would suggest that White players failing to reciprocate would be a more motivationally salient event. However, the opposite pattern would be expected when partners shared (i.e., P3 would be greater for Black than White partners), suggesting that it is more motivationally salient when Black players reciprocate.

During the Trust Game task, the only information participants are provided with to make decisions is a picture of the other player's face. Previous research has demonstrated the significance of the cue stimulus in neural representations of expectancy during reward learning paradigms (e.g., Tortosa et al., 2013). In the first of several exploratory analyses of P2, I examined whether P2 differed by race, and whether this effect was moderated by individual differences in racial bias. The P2 race effect has been demonstrated to be robust to the influences of external factors, such as type of task (e.g., Ito & Urland, 2003, 2005; Correll, Urland, & Ito, 2006). Thus I expected to find that P2

is greater for outgroup (Black) than ingroup (White) faces among White participants, consistent with previous literature. Correll and colleagues (2006) found that the P2 race effect was linked to the extent to which one endorses stereotypes about Black Americans being more violent, as well as bias observed during an implicit bias task. In an effort to examine these effects in the present data, I investigated whether individual differences in racial bias moderate the effect of race on P2. I expected to find that the P2 race effect (i.e., the Black-White difference in P2 signal) would be larger for individuals who hold relatively negative attitudes towards Black people.

One additional exploratory question I hoped to address with this experiment is whether neural signals associated with initial perceptual face processing (i.e., P2) predict subsequent judgments and behavior. Specifically, can P2 elicited by the other player's face predict the amount of money the participant offers to the other player? This was addressed in exploratory analyses.

Method

Participants

Seventy-eight (41 females) undergraduate students enrolled in the Introductory Psychology course at the University of Missouri voluntarily participated in this experiment in exchange for research credit. Of the sample, 1% were Asian, 1% Biracial, 8% Black, 3% Hispanic, and 87% White. I had aimed to have a minimum of 60 participants with valid data, which is slightly larger than samples of 50 obtained in previous, comparable studies (Stanley et al., 2011; Kubota et al., 2013).

Analyses of P2 amplitude values excluded non-White participants because P2 is sensitive to ingroup/outgroup membership status (and, unfortunately, there was an insufficient number of Black participants to split the sample by race). Following these exclusions, 67 participants (33 females) remained in the sample.

Exclusion Criteria Considerations

Originally, I had planned to exclude any participants who did not believe the database stimulus manipulation. However, almost all participants believed that the other players in the game actually came from a database. During the debriefing process, the experimenter found that only two participants had doubts about whether the feedback they received during the TG was from real people (and not randomly generated by the computer). Because these participants were not certain about their doubts, I decided to include them in analyses. Rather than excluding their data, I generated a “global suspicion” variable based on participants’ responses given during the Post-Experiment Questionnaire and the debriefing interview. This variable was included in the analysis of the behavioral responses to determine whether it affected offer amounts.

Non-White participants were included in most analyses (except for those involving P2, where race of the participant would change the predicted outcomes). The focus of the present experiment is on stereotype-dependent social expectancies ostensibly derived from the race of the other player. Anyone is susceptible to making decisions based on racial stereotypes, regardless of their own race or ethnicity (see Kubota et al., 2013). Additionally, inclusion of non-White participants was also important for analyses examining individual differences, because to do otherwise could potentially restrict the range of responses.

General Procedure

Upon entering the lab, participants were consented and informed that the purpose of the present study was to investigate neural responses during economic decision-making processes within the context of social interactions. Participants then received instructions for how to play the TG. Afterwards, they were told about the Database Project, the deception manipulation aimed at leading them to believe the people they would be playing against were real people who had participated previously. As part of this manipulation, participants were introduced to the objectives of the project, and given an opportunity to participate in it themselves. They were given a second (fake) consent form and a photograph release waiver to read and sign. Additionally, their picture was taken with a digital camera. The EEG cap was applied to the participant's scalp (approximately 20 min), and they were asked to complete the very brief Database Project Task (another element of the deception manipulation). Then participants completed five practice trials of the TG paradigm before beginning the task. The non-practice TG trials were divided into four blocks separated by self-timed breaks. The full TG task duration

was approximately 30 minutes. The EEG cap was removed and participants were taken to the bathroom to clean up, prior to completing the post-experiment questionnaires (which took approximately 10 min). Then two trials were randomly selected, and participants were paid immediately in cash. Finally, participants were given a funneled debriefing to probe for suspicion about the deception manipulation and their awareness and thoughts about other aspects of the experiment.

Trust Game Paradigm

Participants played a series of “one-shot trust games” – that is, the other player was different during each trial, and participants never played the same player twice.

Instructions. Participants were told they would be playing the role of the investor (which the experimenter referred to as “Player A”) while playing against many different players in the role of the trustee (referred to as “Player B”). At the beginning of each trial, participants automatically received \$5 to divide (in whole-number increments from \$1 to \$5) between themselves and the other player. During the instructions, they were told that,

“By sharing, you will possibly be able to make more money than you would have if you didn’t share. However, there is also a possibility that you will lose the money you do share. Your goal throughout the game is to try to make as much money as possible.”

They were told they would be given a picture of the other player in order to give them an idea of who the other player is (Kubota et al., 2013; Stanley et al., 2012, 2011) and that they would see a new person during every trial. The experimenter told participants that,

“Whatever amount of money you decide to share will then automatically quadruple before it reaches the other player. For example, if you choose to share

\$2, it will turn into \$8. If you choose to share \$4, it will turn into \$16, and so on.

Then, the other player has a chance to decide what they would like to do with the quadrupled amount. They can either choose to share half of it back with you, or they can choose to keep all of it.”

Then the experimenter walked them through an example to help illustrate the game, and asked them if they had any questions.

Additionally, participants were told they would only have 4 sec to respond once the face appeared (see Figure 1). Participants were also encouraged to begin thinking about their decision as soon as the face appeared in order to take advantage of the full 4 sec response deadline (instead of waiting for the response screen to begin considering their decision; see Figure 1). They were told the purpose of the response time limit was to encourage them to not spend much time thinking about their decisions (e.g., Kubota et al., 2013; Stanley et al., 2012). Participants were further told that “Often times, people are able to form more accurate impressions of people if they rely more on their ‘gut’” (Kubota et al., 2013; Stanley et al., 2011, 2012).

In addition to receiving course credit for participation, participants were told they would receive bonus money based on the money they earned during two randomly selected trials. Because two trials would come to fruition, participants were instructed (and incentivized) to treat each trial as if it was real. The amount they were actually given was ultimately capped at \$5, but participants were unaware of this cap until the end of the experiment when they received their payment.

Trial structure. The trial structure and timing are illustrated in Figure 1. Each trial began with the presentation of a picture of the other player, accompanied by text

above that read “Player B:” (displayed for 2 sec). The face display appeared separately from the response screen in order to facilitate cleaner data acquisition for face processing ERPs. In the next slide, participants were asked “How much would you like to share with this partner?” and they were allowed to select from the options listed at the bottom of the slide (i.e., \$1-5). Participants were given 2 sec from the display onset to respond via a button box with one of five buttons (corresponding to \$1, \$2, \$3, \$4, and \$5). Their selection was confirmed by outlining their choice with a blue rectangle on the display for the duration of what remained of the 2 sec. This was followed by a fixation cross displayed onscreen (1, 1.2, or 1.4 sec). Participants then received feedback indicating whether the other player had decided to share half of the quadrupled amount back with them (displaying “+ \$” written in green), or keep all of it (“- \$” written in red). If participants failed to respond within the 4-sec deadline, the feedback displayed read “Too slow!” The feedback display did not include the exact numeric amounts won or lost (e.g., + \$4) in an effort to reduce saccade eye movement (e.g., Tortosa et al., 2013). Trials were separated by an inter-trial interval (ITI) displaying a fixation cross for 1, 1.2, or 1.4 sec.

In addition to five practice trials, participants completed 210 trials total in random order. The other player was Black during 80 trials, White during another 80 trials, and for 50 trials the other player was another race (e.g., Asian). Both feedback outcomes (keep, share) occurred 50% of the time within each race condition.

Face stimuli. Face stimuli consisted of 210 color pictures of males with neutral expressions (80 Black, 80 White, and 50 other-race faces, such as Asian). Consistent with previous studies (Kubota et al., 2013; Stanley et al., 2011), all other players were male in order to isolate the effect of race without having to consider the intersection of race and

gender (which is beyond the aims of the present study). Although I was primarily interested in comparing responses to Black and White players, faces of other races were included as fillers in order to obscure the research hypothesis (Kubota et al., 2013); the other-race condition was not intended for analyses. The pictures were selected from a collection of 291 faces culled from different face databases by Stanley et al. (2011). Pictures were selected based on initial pilot testing, in which 44 participants (who were not in the present study) rated the trustworthiness of each of the 291 faces on a scale of 1 (*Extremely untrustworthy*) to 9 (*Extremely trustworthy*). Picture selection sought to maintain the trust ratings distributions within racial categories of the larger sample ($M_s = 4.46, 5.15, \text{ and } 4.87$ for Black, White, and other, respectively).

Stimuli Database Project Deception

Participants were told a long and detailed backstory in order to make them believe that the people they were playing against were real people and elicit more realistic behavior from participants during the TG (see Appendix A for instructions script). They were told that the pictures of other players were in fact real people who had participated previously as part of a massive, ongoing, “real-response stimulus database-building project”. It was explained that project has been taking place as part of a collaboration across multiple labs at different universities in Missouri, including a historically Black college in Jefferson City (this latter detail was to help implicitly justify why and how the sample of other players was so much more diverse than the general population of Missouri). They were told the other players’ responses in the role of Player B (i.e., the trustee) had been recorded previously, and that they would receive feedback based on the real responses given by those people. It was emphasized that when participants made an

offer as Player A (the investor) during the TG, the computer would be giving them feedback based on the real response made by that person, and that this response was specific to the particular amount the participant offered.

In order to further bolster this story, participants were given the opportunity to become part of this stimuli database project (SDP) as well (Stanley et al., 2011; 2012; Kubota et al., 2013). It was reasoned that, because the present investigation sought to measure their responses in the role of Player A (the investor), the experimenter was only interested in collecting their responses as Player B (the trustee) for the purpose of the database.

They were asked to complete a (fake) consent form and photograph release form (authorizing the use of their photograph) to participate (Appendix B). When participants opted to not participate in the SDP, they were still asked to participate in the subsequent procedures so that all participants would have the same experience prior to beginning the experiment. Their picture was taken using a digital camera against a white wall while they maintained a neutral facial expression.

After the EEG cap was applied, they completed a very brief computerized task, during which they responded as Player B to offers from a hypothetical Player A (Kubota et al., 2013). During the instructions, the experimenter emphasized the importance of treating these decisions as if they were real – that is, as if real money were at stake. Participants were presented with a question (*“Player A has shared \$X with you. You now have $\$(X*4)$. What would you like to do?”*) for each offer amount, X (\$1, \$2, \$3, \$4, and \$5) in random order (Stanley et al., 2012). Underneath each question, participants could

select one of two options: “*Share half (i.e., $\$X*2$) with Player A, and keep $\$(X*2)$ ” and “*Keep all ($\$X*4$)*”.*

Questionnaires

After completing the TG task, participants were first given a post-experiment questionnaire (made in-house; see Appendix C) that asked them generally about their experience during the study (“*Did you use any strategies or ‘rules of thumb’ to help you make decisions about how much money to share during the Trust Game?*”) and an initial probing for suspicion (e.g., “*What did you think this study was about?*”). This questionnaire was given on paper so that the experimenter could collect and review it while the participants completed the remaining questionnaires. If any important responses that serve as manipulation checks were unclear or indicated that participants were suspicious, the experimenter brought it up during the funneled debriefing interview.

Demographics. Next, participants were given a demographics questionnaire. They were asked about their gender, age, ethnicity, race (i.e., Hispanic or Non-Hispanic), and country of origin.

Racial bias measures. Individual differences in attitudes and motivations concerning race were measured using two questionnaires. The External/Internal Motivation to Control Prejudice Scales (EMS and IMS for external and internal scales, respectively; Plant & Devine, 1998) measure the extent to which participants are externally and internally motivated to control biased behavior. Specifically, the IMS scale measures the extent to which it is important to an individual for personal reasons to attempt to be unbiased toward Black people. The EMS scale measures the extent to which people attempt to control the expression of bias due to external pressures, such as

not wanting to appear bigoted. Participants responded to 10 items (e.g., *“I try to hide any negative thoughts about Black people in order to avoid negative reactions from others”* for EMS; *“Being non-prejudiced toward Black people is important to my self-concept”* for IMS) on a scale from 1 (*Strongly disagree*) to 9 (*Strongly agree*). IMS and EMS are both scored by averaging across the five items specific to each scale (after reverse scoring some questions). Higher scores indicate a greater level of internal or external motivation to control biases.

Participants also completed the Attitudes toward Blacks (ATB; Brigham, 1993) questionnaire. The ATB contains 20 items (e.g., *“I would rather not have Black people live in the same apartment building I live in.”*), each rated on a scale of 1 (*Strongly disagree*) to 7 (*Strongly agree*). The overall ATB score is computed by averaging across the items (after some items have been reverse scored). Higher ATB scores correspond to more negative feelings toward Black people.

Funneled Debriefing

At the end of the experiment, participants were debriefed via a funneled debriefing interview process in order to probe for suspicion (see Appendix D for general outline and script). The experimenter began by asking broad, open-ended questions about the participant’s experience during the experiment (e.g., *“Did any of the tasks seem odd and make you question why you were being asked to do them?”*) instead of asking leading questions (e.g., *“Did you believe the backstory about the stimuli database project?”*).

Data Reduction and Cleaning

Individual differences in racial bias. One of the primary aims of the present study was to investigate whether individual differences in racial bias moderated behavioral and neural race differences observed during the TG. Previous research has shown that individual differences in implicit racial bias (measured via the implicit association test; IAT; e.g., Greenwald et al., 1998) is associated with racially biased behavior during behavioral economic games with diverse partners (Kubota et al., 2013; Stanley et al., 2011, cf. Stanley et al., 2012). Stanley et al. (2011) also tested the relationship between more direct, explicit measures of anti-Black bias (including IMS and EMS) and TG behavior. However, the authors were solely interested in demonstrating that the IAT independently predicted bias during the TG above and beyond explicit measures. The results reported suggest that EMS (but not IMS) predicted preferential treatment of White players during the TG. Although the present study assessed multiple constructs related to racial bias (i.e., biased attitudes with the ATB, and internal/external motivation to control bias with IMS/EMS), a comparison of these measures was not intended.

The only goal in this study was to obtain a general measure of individual differences in racial bias. Initial considerations were given to creating an overall measure of racial bias by combining ATB, IMS, and EMS. This would have been done by creating a latent variable of racial bias using a confirmatory factor analysis (CFA). Outliers in all three measures were handled by winsorizing values that were 3 *SDs* above/below the median score, and Pearson correlations among the variables were examined. ATB and IMS were highly correlated ($r = -.73$); however, correlations involving EMS were much lower ($r_s = .24$ and $-.05$ for ATB and IMS, respectively). These low correlations

indicated that EMS was a distinct construct, and that combining these three variables was not appropriate. I decided to instead focus on the ATB scale, because the construct it measures (i.e., anti-Black sentiments) is most directly related to the general concept of racial bias. Both IMS and EMS were dropped from further analyses. ATB scores were grand mean-centered.

Global suspicion variable. I was concerned that participants might respond differently during the trust game if they a) were not convinced that they were not actually playing to receive real money at the end of the experiment¹, b) did not believe that the feedback they received represented real responses given previously by real people, or c) were aware of my hypotheses about race affecting behavior. Separate suspicion variables were created for each of these three concerns using responses obtained during the post-experiment questionnaire and funneled debriefing (for more details, see Supplemental Material). Participants were rated separately along each of these three dimensions, following similar general guidelines: participants who did not express any suspicion were coded as 0, uncertain or implied suspicion was coded as 1, and more certain or directly expressed suspicion was coded as 2 (similar to Rubin & Moore, 1971; the coding distribution across all three variables was as follows: 33 were coded as 0, 24 as 1, and 21 as 2). To create a global suspicion variable, these three variables were summed, and sums equal to or greater than 2 were recoded as 2 (to avoid influential outliers).

¹ I have found that this can be an issue in the past when participants enroll in the experiment expecting to only receive class research credits as payment.

Electrophysiological recording and processing. The electroencephalogram (EEG) was recorded using an array of 34 Ag/AgCl electrodes fixed in a stretch-lycra cap (ElectroCap, Eaton, OH) placed on the scalp in standard locations (American Encephalographic Society, 1994). Scalp electrodes were referenced online to the right mastoid and an average mastoid reference was calculated offline. Vertical and horizontal electrooculographic activity was recorded with additional electrodes placed 1 cm above and below the left eye and approximately 2 cm outside the outer canthus of each eye, respectively. An additional electrode was placed on the tip of the nose.² Conductive gel was inserted in each electrode to reduce impedance (kept below 8 k Ω at all locations). EEG signals were amplified with Synamps2 amplifiers (Compumedics-Neuroscan, Charlotte, NC), sampled at 500 Hz for the first five participants and 1000 Hz for the rest, and filtered online at 0.01-40 Hz. The sampling rate of 1000 Hz for 73 subjects was converted to 500 Hz so that they could be analyzed together with the first five participants. Ocular artifacts were removed offline using a regression-based procedure (Semlitsch, Anderer, Schuster, & Presslich, 1986). Feedback ERPs (P3 and RewP) were baseline-corrected to 200 - 0 ms pre-feedback stimulus, and P2 ERPs were baseline-corrected to 100 - 0 ms pre-face stimulus. Following rejection of trials containing movement and other voltage artifacts, a bandpass filter of .1 to 20 Hz was applied to the feedback-locked waveforms.

² This electrode was included with the intention of using it as a reference for the N170, an additional ERP component known to be involved in face processing. However, the experimenters had difficulty getting a low impedance from the nose electrode. As a result, the data collected for N170 was not of sufficient quality for analysis.

Visual representations of the waveform were created by first averaging the ERP signal within each race condition for each subject, and then aggregating across subjects to create an overall mean waveform per race condition (i.e., grand average waveforms). Electrodes included in analyses were a swath of electrodes surrounding the maximal peak of the waveform of interest. Electrode selection was determined by visual inspection of the grand average waveforms in conjunction with taking into account what previous studies have done (see Figure 2). Analyses for RewP included 15 electrodes centered at the fronto-central midline scalp location (i.e., F1, F2, F3, F4, Fz, FC1, FC2, FC3, FC4, FCz, C1, C2, C3, C4, and Cz). Analyses for P3 and P2 included 15 electrodes centered at central parietal and posterior midline scalp locations (C1, C2, C3, C4, Cz, CP1, CP2, CP3, CP4, CPz, P1, P2, P3, P4, and Pz). Each ERP component of interest was quantified for each trial by averaging the signal within a specified interval that included the waveform (RewP: 230-330 post-feedback; P3: 330-400 ms post-feedback; P2: 135-185 ms post-face stimulus).

Analytic Approach

Within-subject predictors were mean-centered within subjects in order to remove all between-person variation (i.e., centering within clusters, or CWC, as discussed by Enders & Tofghi, 2007). Between-subject variables that lacked a meaningful zero were centered to the grand mean (CGM).

For all models involving data collected at multiple levels, I modeled the trial level data using multilevel models (MLM). In addition to accommodating differing patterns of missing observations across individuals, MLM has several other advantages over univariate repeated-measures ANOVA for analyzing psychophysiological data (for in-

depth consideration of these issues, see Bagiella, Sloan, & Heitjan, 2000; Kristjansson, Kircher, & Webb, 2007; Page-Gould, 2017). MLMs of TG response data contained 2 levels (trial and subject) and contained random effects for trial (level 1), face picture stimulus (level 1), and subject (level 2). MLMs of ERPs (units in μV) were 3-level models with random effects of electrode (level 1), trial and face stimulus (level 2), and subject (level 3). MLMs were analyzed in R version 3.4.3 (R Core Team, 2017) with the lmer function that is part of the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Decisions to retain random slopes were made separately for each MLM using the bottom-up approach described by Matuschek and colleagues (2017) that requires starting with the maximal model, and removing highly correlated random slopes (with a cutoff of .8; see Table S1 in Supplemental Material for random effects included in each MLM). An unstructured covariance structure was assumed and degrees of freedom were estimated using the Satterthwaite method. Significance of the F -tests were computed for each effect using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Effect sizes for effects were calculated using the partial R^2 method (Edwards, Muller, Wolfinger, Qadish, & Schabenberger, 2008), where $R^2 < .1304$ is considered a small effect, $.1304 \leq R^2 < .2592$ is a medium effect, and $R^2 \geq .2592$ is a large effect (Cohen, 1992).

Interactions involving continuous variables were probed using the approach outlined by Aiken and West (1991), where the simple slopes are tested at 1 SD above and below the mean of the continuous variable. When degrees of freedom were above 200 for t -tests, the z statistic was reported instead. All simple effect means for levels of the categorical variables reported here are the model-specific estimated marginal means (obtained using the lsmeansLT function in the lsmeans R package; Lenth, 2016). For

estimated marginal means, 95% CIs are reported. All graphs were created using ggplot2 (Wickham, 2009). The effects R package (Fox, Weisberg, Friendly, & Hong, 2016) was used to extract predicted values and confidence bands displayed in the interaction plots.

Results

Analyses of Primary Hypotheses

Trust Game Offers.

Race. A MLM was used to test whether partner's race predicted offer amounts (at the trial level) during the TG. This analysis showed a marginal effect of race, $F(1, 130.58) = 2.97, p = .087, R^2_{\text{partial}} = .02$. The amount of money participants offered to White and Black partners during the TG did not differ significantly ($M_s = \$2.76$ and $\$2.90, SE_s = 0.09$ and $0.07, CIs [2.58, 2.93], [2.75, 3.04]$, for Black and White partners, respectively).

Although it was feasible that non-White participants could exhibit biased behavior during the Trust Game, it was important to test this effect in the sample of White participants because of the longstanding history of racism among White Americans.³ In a separate MLM including data from only the White participants ($n = 67$; full sample $n = 78$)⁴, the effect of race was significant, $F(1, 129.49) = 6.10, p = .015, R^2_{\text{partial}} = .05$. White participants tended to share slightly more (i.e., $\$0.19$) with White partners ($M = 2.88, SE = 0.08, CI [2.72, 3.03]$) than with Black partners ($M = 2.69, SE = 0.90, CI [2.51, 2.87]$). Although the effect of race was “significant” in the White participants and “not

³ It would have been ideal to include participant's race as a moderator; however, there were only 11 participants who were not White. The mean difference in offers made to Black and White partners (i.e., mean offers made to White partners subtracted from the mean offers made to Black partners) for all non-White participants was $\$0.16 (SD = 1.04; \text{range: } -1.29 \text{ to } 2.83)$, for Black participants ($n = 6$) was $\$0.31 (SD = 1.27; \text{range: } -1.29 \text{ to } 2.83)$, and for White participants it was $-\$0.19 (SD = 0.51; \text{range: } -1.85 \text{ to } 0.71)$.

⁴ Although analyses with only White subjects reduces the sample size, it should be noted that it is still above the *a priori* minimum sample size goal of 60.

significant” in the full sample, the effect only changed slightly from the analysis with the full sample (i.e., the change in mean race difference was only \$0.05).

Global Suspicion. I tested whether participants’ level of suspicion about the manipulations affected their behavior during the trust game. Suspecting that the other players weren’t real, thinking that there was no possibility of receiving money at the end of the experiment, or being aware of the race hypotheses might have led participants to make riskier decisions (e.g., sharing more money with people they would have otherwise judged as untrustworthy). Given this reasoning, individuals with higher suspicion scores might have shared more with Black players than White players. To test this, I conducted a Race x Suspicion MLM of offer amount. Adding suspicion and the interaction term did not change the results for the main effect of race, $F(1, 99.80) = 3.11, p = .081, R^2_{\text{partial}} = .03$. Effects of Suspicion and the Race x Suspicion interaction were not significant, $F_s < 2, p_s > .15$. This reveals no evidence that level of suspicion systematically impacted offers made during the TG.

Race x Attitudes toward Blacks (ATB). To test whether individual differences in negative attitudes toward Black people predicted differential treatment of Black and White partners during the TG, a 2 (Race) x ATB (higher scores denote more negative attitudes) MLM of offer amount was conducted. Main effects of race, $F(1, 133.88) = 3.19, p = .076, R^2_{\text{partial}} = .02$, and ATB, $F(1, 76.01) = 1.24, p = .268, R^2_{\text{partial}} = .016$, were not significant. There was a significant Race x ATB interaction, $F(1, 76.06) = 9.04, p = .004, R^2_{\text{partial}} = .11$ (Figure 3). Tests of the simple slopes indicated that the slope of race (White = -1, Black = 1) was significant for higher ATB scores (i.e., more biased individuals; $b = -0.17, SE = 0.05, t(106.05) = -3.31, p = .001$, but not lower ATB scores,

(i.e., less biased individuals; $b = 0.03$, $SE = 0.05$), $t(106.03) = 0.59$, $p = .556$, suggesting that individuals with greater bias shared more money with White than Black partners. Additionally, the ATB slope was significant for Black partner trials ($b = -0.21$, $SE = 0.10$), $t(80.0) = -2.29$, $p = .032$, but not White partner trials ($b = 0.03$, $SE = 0.08$), $t(76.03) = .38$, $p = .703$. This means that higher ATB scores (indicating greater race bias) were associated with sharing less money with Black partners, whereas the amount shared with White partners was not related to ATB scores.

Reward Positivity (RewP). Recent work has emphasized the importance of examining the difference between wins and losses in the RewP signal (e.g., Proudfit, 2015). Because I was interested in analyzing the data at the trial level, I was unable to examine the difference waveform (because each trial can only be either a win or a loss). For this reason, I included interactions involving feedback type as a predictor in all models as a way of testing whether any of these variables moderated the difference between wins and losses.

Race x Feedback. To test the prediction that there would be a main effect of race on RewP, I conducted a 2 (Race) x 2 (Feedback; Share, Keep) MLM of the quantified RewP signal at the trial level. There was a significant main effect of feedback (Figure 4A), $F(1, 142.99) = 51.24$, $p < .001$, $R^2_{\text{partial}} = .26$, confirming the typical RewP finding that RewP is greater for wins (i.e., when the partner shared; $M = 8.89$, $SE = 0.64$, CI [7.61, 10.17]) than losses (i.e., when the partner kept the money; $M = 6.45$, $SE = 0.61$, CI [5.22, 7.67]). However, contrary to predictions, the main effect of race was not significant (Figure 4A), $F(1, 350.30) = 0.20$, $p = .654$, $R^2_{\text{partial}} < .001$, meaning there is no evidence that RewP differed in response to feedback about the decisions made by Black

and White partners ($M_s = 7.58$ and 7.74 , $SE_s = 0.62$ and 0.64 , $CI_s [6.37, 8.82]$ and $[6.48, 9.01]$ for Black and White, respectively). When considering RewP as a measure of expectancy, this result suggests that expectancies of Black and White partners did not differ. Additionally, the Race x Feedback interaction was not significant, $F(1, 77.08) = 0.15$, $p = .704$, $R^2_{\text{partial}} = .002$.

Race x Feedback x Offer Amount. Because RewP can fluctuate as the result of differences in magnitude of monetary gains and losses (e.g., Kreussel et al., 2012), it was important to include offer amount in the model. Additionally, because level of trust (and expectations about the outcome) is implied in offer amount, I was interested to see if racial differences in RewP emerged for feedback when expectations were higher or lower. To test this hypothesis, I ran a 2 (Race) x 2 (Feedback) x Offer (mean-centered within subjects) MLM of RewP at the trial level. In addition to the main effect of feedback found in the previous model, $F(1, 75.97) = 60.66$, $p < .001$, $R^2_{\text{partial}} = .44$, there was also a main effect of offer amount, $F(1, 63.09) = 42.78$, $p < .001$, $R^2_{\text{partial}} = .40$. A positive slope ($b = 0.97$, $SE = 0.15$) indicates that RewP was greater in response to feedback pertaining to higher offers.

The Race x Offer interaction was also significant, $F(1, 62.41) = 5.61$, $p = .02$, $R^2_{\text{partial}} = .08$ (Figure 5). Simple slope tests indicate there was a significant positive effect of race (coded as White = -1, Black = 1) for higher offers, ($b = 0.42$ $SE = 0.20$), $t(104) = 2.10$, $p = .039$, and a nonsignificant negative slope of race for lower offers ($b = -0.32$, $SE = 0.19$), $t(100.71) = -1.63$, $p = .105$. This suggests that on trials when participants made higher offers, RewP was greater for Black than White partners; however, for lower offers, RewP did not differ by partner's race. The direction of the former effect is consistent with

my original prediction that RewP would be greater for Black than White partner trials. Additional analyses revealed the effect of offer amount was significant for both Black partner trials ($b = 1.27$, $SE = 0.22$), $t(63.18) = 5.67$, $p < .001$, and White partners ($b = 0.66$, $SE = 0.186$), $t(74.03) = 4.07$, $p < .001$, although the effect is notably larger for the former (see Figure 5). This suggests the relationship between offer amount and RewP is stronger for Black partners than White partners. All other effects in the model were not significant, $F_s < 1$, $p_s > .6$.

Race x Feedback x ATB. I expected to find that the effect of race on RewP amplitude would be moderated by individual differences in racial bias (measured by ATB). Specifically, I predicted that the race effect on RewP would be more pronounced for individuals with more negative attitudes toward Blacks (ATB). I tested this hypothesis in a 2 (Race) x 2 (Feedback) x ATB MLM on trial-level RewP data. As before, there was a significant effect of feedback, $F(1, 76.20) = 68.75$, $p < .001$, $R^2_{\text{partial}} = 0.47$. Additionally, there was also a significant interaction of Race x ATB, $F(1, 76.23) = 12.97$, $p < .001$, $R^2_{\text{partial}} = .15$ (Figure 6). Simple slope tests indicated that for individuals with higher ATB scores (i.e. more bias), RewP was greater when feedback received was from White compared to Black partners ($b = -0.38$, $SE = 0.15$; with race coded as -1 for White and 1 for Black), $t(161.06) = -2.44$, $p = .016$. The direction of this effect is contrary to predictions, as it would suggest that more biased individuals expected White partners to keep the money, and Black partners to share. There was a nonsignificant race effect in the opposite direction observed for participants with lower ATB scores (i.e., less bias); $b = 0.25$, $SE = 0.15$), $t(160.76) = 1.62$, $p = .108$. The ATB slope was not significant for both levels of race, $|ts| < 1.1$, $p_s > .2$.

Exploratory Analyses of Secondary Hypotheses

P3 to Feedback. The same models applied to RewP were also applied to P3 to investigate how the manipulations affected motivational salience.

Race x Feedback. It was predicted that White (relative to Black) partners failing to reciprocate would be a more motivationally salient event (resulting in a greater P3), because such outcomes might be more indicative of updating expectancies and modifying behavior accordingly. Likewise, the opposite pattern was expected when partners reciprocated for the same reasons. This was assessed using a 2 (Race) x 2 (Feedback) MLM of P3 at the trial level. This model yielded a significant main effect of feedback, $F(1, 227.66) = 13.47, p < .001, R^2_{\text{partial}} = .06$, reflecting a greater P3 for feedback indicating money was won ($M = 14.07, SE = 0.75, CI [12.60, 15.60]$) compared to lost ($M = 12.34, SE = 0.75, CI [10.90, 13.80]$) (Figure 4B). Like RewP, the main effect of race and the Race x Feedback interaction were not significant, $F_s < 1, p_s > .5$ (see Figure 4B).

Race x Feedback x Offer Amount. To test whether offer amount moderated the Race x Feedback effect on P3, we conducted a 2 (Race) x 2 (Feedback) x Offer Amount MLM of P3. This model produced results consistent with the base Race x Feedback model above, including a significant main effect for feedback, $F(1, 75.67) = 24.68, p < .001, R^2_{\text{partial}} = .25$, and non-significant effects of race and Feedback x Race, $F_s < 1, p_s > .3$. As expected, there was also a significant main effect of offer amount ($b = 1.87, SE = 0.39, F(1, 60.38) = 47.62, p < .001, R^2_{\text{partial}} = .44$, meaning that P3 was greater when feedback denoted larger wins and losses. All other effects were not significant, $F_s < 2.5, p > .12$.

Race x Feedback x ATB. The Race x Feedback MLM (above) failed to find evidence that Black partners reciprocating and White partners not reciprocating are more motivationally salient. It is possible that this effect only occurs in individual differences who are more biased. This was tested with a 2 (Race) x 2 (Feedback) x ATB MLM of P3. Again, there was a significant main effect of feedback, $F(1, 138.34) = 19.11, p < .001$, $R^2_{\text{partial}} = .12$. Additionally, like RewP, there was a significant interaction of Race x ATB, $F(1, 77.12) = 7.85, p = .006$, $R^2_{\text{partial}} = .09$ (Figure 6). Simple slopes tests indicate that the ATB effect on P3 was not significant for Black ($b = -0.53, SE = 0.82$) or White partner trials ($b = 0.44, SE = 0.88$), $|ts| < 1, ps > .5$. For individuals with higher ATB scores (i.e., more bias), P3 was greater during White than Black partner trials (White coded as -1 and Black as 1; $b = -0.53, SE = 0.24$), $t(136.36) = -2.26, p = .025$ (see Figure 6). This would suggest that for more biased individuals, it is more motivationally salient to receive feedback (regardless of valence) from White partners than Black partners. However, for those with lower ATB scores (i.e., less bias), there was no significant race difference in P3 ($b = 0.26, SE = 0.24$), $t(135.66) = 1.12, p = .266$, suggesting motivational salience did not differ as a function of partner's race.

P2 in Response to Faces. Because social group membership of the participant determines the predicted direction of the race effect on P2 amplitude, all analyses involving P2 were restricted to White participants only ($n = 67$).

Race. I first sought to replicate previous findings showing that P2 is greater in response to faces of outgroup members (i.e., Black partners) compared to ingroup members (i.e., White partners). A MLM of trial-level P2 amplitude with partner's race as a fixed effect revealed a significant effect of race, $F(1, 196.15) = 13.34, p < .001$, R^2_{partial}

= 0.06. Consistent with previous research, the P2 elicited by Black faces ($M = 5.53$, $SE = 0.50$, $CI [4.53, 6.52]$) was greater than the P2 elicited by White faces ($M = 4.37$, $SE = 0.49$, $CI [3.39, 5.34]$) in the sample of White participants (Figure 4C).

Race x ATB. A 2 (Race) x ATB MLM of trial level P2 was analyzed to investigate whether the race effect (i.e., Black > White) on P2 was augmented in participants who endorse more negative feelings toward Black people. This would suggest more biased individuals were allocating more attention to social categories than those with less bias. With ATB included in the model, the effect of race remained significant, $F(1, 196.27) = 14.76$, $p < .001$, $R^2_{\text{partial}} = 0.07$. The main effect of ATB did not reach significance, $F(1, 65.02) = 3.34$, $p = .072$, $R^2_{\text{partial}} = 0.05$, and the Race x ATB interaction was significant, $F(1, 65.33) = 4.31$, $p = .042$, $R^2_{\text{partial}} = 0.06$ (Figure 7). Simple slope tests for the interaction indicated that the P2 was greater on Black compared to White trials for both individuals who had lower ATB scores ($b = 0.81$, $SE = 0.19$), $t(144.23) = 4.21$, $p < .001$, and those who had higher scores ($b = 0.39$, $SE = 0.18$), $t(160.34) = 2.16$, $p = .032$, although this difference was noticeably smaller for the latter (see Figure 7). This suggests that there was a greater Black-White P2 difference observed in people with less race bias. This is contrary to my prediction that the P2 race effect (i.e., the Black-White difference in P2 signal) would be larger for individuals who hold more negative attitudes toward Blacks. Additional tests show the relationship between ATB and P2 in response to White faces was a significant ($b = 1.17$, $SE = 0.50$), $t(65.05) = 2.34$, $p = .023$, whereas the relationship between ATB and P2 for Black faces did not differ from zero ($b = 0.65$, $SE = 0.53$), $t(65.03) = 1.24$, $p = .220$ (see Figure 7). This suggests

that individuals who have more negative sentiments about Black people have an increased P2 in response to White faces, and no change in P2 in response to Black faces.

P2 predicting trial-level TG offer amounts. I was interested in whether P2 predicted subsequent TG offer amounts at the trial level. Additionally, I was interested in whether between-subject variation in the extent to which P2 differed for White and Black faces predicted differences in offers made to Black and White partners. Following the approach outlined by Enders and Tofghi (2007), P2 was disambiguated into a between-subject race difference variable and a trial-level within-subject variable (averaged across electrodes). The between-subjects race difference P2 variable was quantified by subtracting the mean P2 for White faces from the mean for Black faces for each person. I conducted a Race x P2 (within-subjects) + Race x P2 Race Difference (between-subjects) MLM predicting offer amounts (treated as a continuous variable) at the trial level.⁵ Like the earlier model of race predicting offer amount among White participants, there was again a significant main effect of race, $F(1, 111.10) = 9.46, p = .003, R^2_{\text{partial}} = .078$, indicating offers were higher for White ($M = 2.88, SE = 0.08, CI [2.72, 3.03]$) than Black partners ($M = 2.69, SE = 0.09, CI [2.50, 2.87]$). There was also a marginally significant effect of Race x P2 Race Difference, $F(1, 65.01) = 3.39, p = .070, R^2_{\text{partial}} = .05$ (see Figure 8). For exploratory purposes, I decided to probe the marginal interaction. According to the simple slopes tests, the effect of P2 race difference on offer amount was not significant within Black ($b = -0.005, SE = 0.05, t(64.99) = -0.09, p = .927$), or White partner trials ($b = -0.07, SE = 0.04, t(65.03) = -1.71, p = .092$). However, there was a

⁵ The two P2 variables were uncorrelated ($r < .001$).

significant race effect (White = -1, Black = 1) for participants with smaller P2 race differences ($b = -0.19$, $SE = 0.07$), $t(82.91) = -2.95$, $p = .004$, and an almost significant race effect for participants with greater P2 race differences ($b = -0.08$, $SE = 0.04$), $t(125.11) = -1.97$, $p = .052$. This suggests that the tendency to offer more money to White than Black partners was slightly more pronounced in participants with smaller P2 race differences, which is contrary to predictions.

Discussion

Recent research has demonstrated that the race of other players and one's own racial biases can influence monetary decisions during behavioral economics games (Kubota et al., 2013; Stanley et al., 2011). The present experiment sought to extend this research by examining ERP components associated with expectancy (RewP), reward salience (P3), and social categorization (P2). Participants played a behavioral economics Trust Game (TG), during which participants played with male partners of different races. By sharing money, participants had the opportunity to earn more money than they would have if they didn't share; however, there was also a possibility of losing the money they shared. Participants received feedback indicating the outcome, which (from the participant's perspective) was determined by a decision made by the other player to either reciprocate the participant's generosity by sharing back with the participant or keep all of the money for themselves.

Behavior during the Trust Game

I predicted that during the TG participants would be more trusting of – meaning that they would make higher offers to – White players than Black players. In the full sample of 78 participants, the difference in the amount participants shared with Black players and White players (i.e., \$0.14) was not significant ($p = .086$). This is consistent with the other two experiments that found no effect of partner's race on participants' offers during the TG in multiracial participant samples (Stanley et al., 2011; 2012). However, unlike previous experiments (Kubota et al., 2013; Stanley et al., 2011), there was behavioral evidence of bias in the subsample of only White participants. In the present experiment, White participants shared significantly more money (i.e., \$0.19) with

White partners than Black partners. Previous studies' failure to find evidence of preferential treatment within subsamples of White participants may have been due to smaller sample sizes ($n = 67$ White subjects in the present study, compared to 27 and 31 White participants in Kubota et al., 2013 and Stanley et al., 2011, respectively).

Additionally, previous research has found that racial bias exhibited during behavioral economic games with racially diverse partners is moderated by measures of implicit (i.e., indirect) racial bias measured via the IAT (Kubota et al., 2013; Stanley et al., 2011) and an explicit (i.e., direct) racial bias (Stanley et al., 2011). I sought to conceptually replicate these findings by demonstrating that this moderation effect could be obtained by using a different explicit measure – specifically a validated questionnaire that measures negative attitudes toward Black people (i.e., the ATB). Consistent with predictions, individuals with more negative attitudes toward Black people shared more money with White than Black partners, whereas in individuals with less negative attitudes, there was no evidence of offer amount varying for Black and White partners. Additionally, having a more negative attitude toward Black people was associated with sharing less money with Black partners, whereas the amount shared with White partners was not related to the attitudes measure. This replicates the relationship found by Stanley et al. (2011) with the IAT.

ERP Correlates of Expectancy (RewP) and Motivational Salience (P3)

The present experiment was designed to examine the ERP component, RewP, in response to feedback during the TG. RewP was of interest because of its associations with social expectancies during social behavioral economics games (e.g., Chen et al., 2012). A second ERP component known as P3 is also elicited by performance feedback,

and is often examined in addition to RewP in reward learning experiments. P3 is thought to reflect motivational salience (Nieuwenhuis et al., 2005). Analyses of P3 were all exploratory; however, since the results obtained in the RewP and P3 analyses were so similar, it is simpler to reviewed and discuss them together.

First, I replicated commonly found effects of P3 and RewP that were unrelated to my hypotheses regarding race. Consistent with previous literature on RewP and P3, I found robust effects of feedback, where RewP and P3 were greater for gains (i.e., partner shared back) than losses (i.e., partner kept the money). However, the magnitude of gains was higher than the magnitude of losses during the TG, which could have contributed to this finding for P3. Additionally, both RewP and P3 increased with the magnitude of the offers (regardless of feedback valence), which also replicates previous findings (e.g., Kreussel et al., 2012; Sato et al., 2005; Yan Wu & Zhou, 2009).

I predicted that the other player's race would inform expectancies during the game, and that this would be reflected in differential amplitudes in the RewP when provided with feedback about the other player's decision. Specifically, in response to both positive and negative feedback, RewP magnitude was predicted to be greater for Black than White partners. For positive feedback, this result would imply that Black partners sharing back with participants was better than expected. For negative feedback, this would suggest that it was more unexpected for White players to keep the money than Black players. However, there was no evidence that RewP differed for Black and White partners across either type of feedback.

Additional analyses indicated that offer amount (within-subjects) interacted with race to predict RewP. On trials when participants made higher than average offers, RewP

was greater for feedback from Black partners than White partners, whereas there is no race difference in RewP for lower offers. The former effect is in the predicted direction. One explanation for this result pertains to the fact that expectancies and perceived level of trust are presumably taken into consideration when a participant decides how much to offer a partner on a given trial. With this in mind, this result may perhaps imply that when more trusted Black partners shared back, this was more unexpected, and when more trusted Black partners kept the money, this was expected. Alternatively, a different explanation is that this RewP effect could simply be occurring when there is more money at stake, and is therefore more salient. The increased salience of higher offers for participants is supported by the very large effect offer amount has on P3. When the stakes are higher, that might be when participants' expectancies about the other players are strong enough to produce the expected race effect in RewP. Regardless of the explanation, because these results were not exactly in line with my predictions, more research should be done before conclusions can be drawn.

Additionally, there was a more pronounced Black-White difference in both RewP and P3 for individuals with more negative attitudes toward Black people. However, the direction of the difference was in the opposite direction from what I had predicted. For more biased people, RewP and P3 were greater when feedback received was from White compared to Black partners. However, for less biased participants, RewP and P3 did not differ across race. If we assume expectancies are the sole contributing factor in this pattern observed in RewP, this finding would imply that more biased participants were more surprised when White partners shared back, and less surprised when they kept the money. Likewise, the P3 results would suggest that feedback received from White

partners was more motivationally salient than feedback from Black partners. While these interpretations are possible, they both seem reasonably unlikely in light of previous research. For example, (Mendoza et al., 2014) found evidence that ingroup members expect preferential treatment to be reciprocated by fellow ingroup members. For a predominantly White sample like the one at hand, this suggests participants' expectancies should be violated when White partners don't reciprocate. This contradicts the present results.

The more likely interpretation is that these results actually reflect differences in *magnitude*. Because more biased individuals made higher offers to White partners, the feedback they received regarding these offers were on average of a higher magnitude. This is an issue I had not previously considered, and given the large effects that offer amount had on both P3 and RewP (i.e., $R^2_{\text{partial}} = .44$ and $.40$ for P3 and RewP, respectively), this seems to be the most likely explanation for the results on hand. It would be ideal to run an additional analysis controlling for offer amount in order to examine the effect of individual differences in racial bias on both the P3 and RewP results. However, because of concerns of multicollinearity among the predictors (especially ATB, offer amount, and race), I did not control for offer amount.

Exploratory Examination of the ERP Correlate of Social Categorization (P2)

Replicating previous research (e.g., Amodio, 2010; Dickter & Bartholow, 2007; Dickter & Kittel, 2012), I found that P2 was greater for racial outgroup (Black) than ingroup (White) faces in White participants during the TG.

Consistent with previous research (Correll et al., 2006), I predicted that the P2 race effect (Black-White) would be larger for individuals who hold more negative

attitudes toward Black people. However, contrary to predictions, there were greater Black-White P2 differences observed in participants with less negative attitudes toward Black people; this effect was driven by a decrease in P2 to White faces for these individuals (relative to those with more bias). Consistent with this finding, an additional analysis revealed trending evidence that larger Black-White race differences in P2 were associated with less biased behavior exhibited during the TG (i.e., no difference in amount offered to White and Black partners). These results suggest that participants who are less biased (when assessed via the ATB or TG) attended more to the racial categories of their partners compared to more biased individuals. Although this was contrary to expectations, this could have resulted from less biased participants attempting to avoid biased behavior; doing so may have required them to actually attend more to racial category information. Future analyses could be done to address this possibility.

Additionally, the present P2 results seem to contradict previous research by Correll et al. (2006) demonstrating greater P2 signals for Black relative to White faces were associated with more biased task behavior and stronger violent Black stereotype endorsement. However, the task they used to assess biased behavior is a fast-paced first person shooter task, which may measure different underlying constructs of racial bias when compared to the TG (e.g., violence/aggression vs. financial trustworthiness). Additionally, violent stereotypes endorsement is quite a different construct from the anti-Black sentiments measured with the ATB. These methodological differences may explain the opposing results obtained in the present study. Ultimately, future research should seek to replicate the present effect, since it is not supported by previous literature.

Additionally, I examined whether within-subject variation in P2 predicted subsequent offer amounts at the trial level. However, I did not find evidence that P2 in response to faces predicted subsequent behavior (i.e., offer amount) during the TG at the trial level.

Limitations and Future Directions

One limitation of the present research is that there was an insufficient number of non-White participants to analyze how participant's race impacted behavior during the TG. Theoretically speaking, the tendency to not trust Black males with money does not necessarily exclusively occur in White people. However, it appears that this assumption was not supported by the present data. There was actually a stronger effect of race bias in the subsample of White participants. The unbalanced numbers of participants belonging to different racial categories prevented me from making a meaningful comparison of participant's race. However, inspection of the mean partner race difference in offer amounts for non-White participants (especially Black participants) provides preliminary evidence that participant's race and target's (i.e., partner) race may interact to affect behavior during the TG. Additional evidence that behavior during this game most likely varies across races comes from the present study's finding that biased behavior during the TG is related to explicitly expressed racially biased attitudes – such attitudes can be expected to vary across different races. Future research should investigate and compare the extent and direction of biases evident in social behavioral economic games (such as the TG) in people of different races.

An additional limitation of the current TG paradigm is that participants knew that their behavior would not affect the other players. The partners represented in the pictures

were never going to see how much the participant had offered them, and there was no possibility for the partners to receive any money as a result of the decisions made by the participant during the TG. In effect, the possibility of harming other players (e.g., causing feelings of disappointment, or depriving them of money) was removed in this version of the trust game. This kind of “harm-free” scenario is rare in society. Even in situations when decisions are made anonymously, they are rarely without consequence. This may have inadvertently created a space in which participants felt free to be biased without the restraints of potentially causing harm. This is to say that individuals who harbor biased beliefs (e.g., negative attitudes toward Black people) might believe that the most financially advantageous strategy is to use race and associated stereotypes to inform their decisions. This possibility is contradicted by the P2 evidence suggesting that more biased people were attending less to racial categories. However, some anecdotal evidence in favor of this possibility comes from the open responses in the post-experiment questionnaire, where several participants readily admitted that they made decisions based on race (among other factors, such as age). This harm-free environment created in the present experiment cannot be ruled out as an explanation for the biased behavior demonstrated by White participants during the present TG task.

Additionally, in an effort to isolate the effect of race, the present experiment only used male faces to represent different race partners. In doing so, the effects found in the present study can only be said to generalize to males. Future research should seek to investigate the intersectionality of gender and race in perceived trustworthiness when money is concerned.

Conclusion

The present experiment tested whether race affects perceived trustworthiness during a behavioral economics game. In the full sample of participants, there was no evidence that race affected behavior. However, there was an effect of race among White participants, who demonstrated a higher level of trust in White players than Black players during the game. Across all participants, this effect was exaggerated in individuals with more racially biased beliefs. Additionally, RewP was greater for feedback from Black relative to White players on trials when offer amounts were higher than average. This implies that when feedback was most salient (because more money was at stake), it was more unexpected when Black players reciprocated generosity, and less expected when they did not. Exploratory analyses of P3 only seemed to reiterate some of the RewP effects. In exploratory analyses of P2, it was observed that greater P2 race differences (Black-White) in less biased individuals suggests that these participants may have allocated more attention to social categories, perhaps as the result of attempting to respond in a less biased way. These novel results were not entirely consistent with hypotheses, and thus more research is required to replicate and help explain these results.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Amodio, D. M. (2010). Coordinated roles of motivation and perception in the regulation of intergroup responses: frontal cortical asymmetry effects on the P2 event-related potential and behavior. *Journal of Cognitive Neuroscience*, 22(11), 2609–2617.
<https://doi.org/10.1162/jocn.2009.21395>
- 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. *Journal of Personality and Social Psychology*, 90(2), 272–287.
<https://doi.org/10.1037/0022-3514.90.2.272>
- 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>
- Bartholow, B. D., Pearson, M. A., Gratton, G., & Fabiani, M. (2003). Effects of alcohol on person perception: A social cognitive neuroscience approach. *Journal of Personality and Social Psychology*, 85(4), 627–638. <https://doi.org/10.1037/0022-3514.85.4.627>
- Bates, B., Douglas, M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>

- Becker, M. P. I., Nitsch, A. M., Miltner, W. H. R., & Straube, T. (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *The Journal of Neuroscience*, *34*(8), 3005–3012. <https://doi.org/10.1523/JNEUROSCI.3684-13.2014>
- Begleiter, H., Porjesz, B., Chou, C. L., & Aunon, J. I. (1983). P3 and stimulus incentive value. *Psychophysiology*, *20*(1), 95–101. <https://doi.org/10.1111/j.1469-8986.1983.tb00909.x>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Boksem, M. A. S., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience*, *5*(1), 118–128. <https://doi.org/10.1080/17470910903202666>
- Brigham, J. C. (1993). College Students' Racial Attitudes. *Journal of Applied Social Psychology*, *23*(23), 1933–1967. <https://doi.org/10.1111/j.1559-1816.1993.tb01074.x>
- Burton, L. M., Mattingly, M., Pedroza, J., & Welsh, W. (2017). Poverty. *Pathways Magazine, Special Issue: State of the Union*, 9–12.
- Campanhã, C., Minati, L., Fregni, F., & Boggio, P. S. (2011). Responding to Unfair Offers Made by a Friend: Neuroelectrical Activity Changes in the Anterior Medial Prefrontal Cortex. *Journal of Neuroscience*, *31*(43), 15569–15574. <https://doi.org/10.1523/JNEUROSCI.1253-11.2011>

- Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. *NeuroImage*, 57(4), 1608–1616. <https://doi.org/10.1016/j.neuroimage.2011.05.037>
- Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2010). Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. *Journal of Cognitive Neuroscience*, 23(4), 936–946. <https://doi.org/10.1162/jocn.2010.21456>
- Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience Letters*, 517(2), 65–70. <https://doi.org/10.1016/j.neulet.2012.02.082>
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314–1329. <https://doi.org/10.1037/0022-3514.83.6.1314>
- 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>
- Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, 2(3), 189–198. <https://doi.org/10.1093/scan/nsm012>

- Dickter, C. L., & Kittel, J. A. (2012). The effect of stereotypical primes on the neural processing of racially ambiguous faces. *Social Neuroscience*, 7(6), 622–631.
<https://doi.org/10.1080/17470919.2012.690345>
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. <https://doi.org/10.1037/1082-989X.12.2.121>
- Fischer, C. S., & Hout, M. (2006). *Century of Difference: How America Changed in the Last One Hundred Years*. New York, NY: Russell Sage Foundation.
- Foti, D., Carlson, J. M., Sauder, C. L., & Proudfit, G. H. (2014). Reward dysfunction in major depression: multimodal neuroimaging evidence for refining the melancholic phenotype. *NeuroImage*, 101, 50–58.
<https://doi.org/10.1016/j.neuroimage.2014.06.058>
- Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, 32(12), 2207–2216. <https://doi.org/10.1002/hbm.21182>
- Fox, J., Weisberg, S., Friendly, M., & Hong, J. (2016). R Package “effects”: Effect displays for linear, generalized linear, and other models (Version 3.1-2).
- Gehring, W. J., & Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*, 295(5563), 2279–2282.
<https://doi.org/10.1126/science.1066893>

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27.

<http://dx.doi.org/10.1037/0033-295X.102.1.4>

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.

<https://doi.org/10.1037/0022-3514.74.6.1464>

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes.

Psychophysiology, 42(2), 161–170. <https://doi.org/10.1111/j.1469-8986.2005.00278.x>

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: the feedback negativity and violations of reward prediction in gambling

tasks. *Psychophysiology*, 44(6), 905–912. <https://doi.org/10.1111/j.1469-8986.2007.00567.x>

He, Y., Johnson, M. K., Dovidio, J. F., & McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Social Neuroscience*, 4(5), 426–442.

<https://doi.org/10.1080/17470910902949184>

- Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G. H., Holroyd, C. B., & Miltner, W. H. R. (2011). Why humans deviate from rational choice. *Psychophysiology*, 48(4), 507–514. <https://doi.org/10.1111/j.1469-8986.2010.01081.x>
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709. <https://doi.org/10.1037/0033-295X.109.4.679>
- Holroyd, C. B., Hajcak, G., & Larsen, J. T. (2006). The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. *Brain Research*, 1105(1), 93–101. <https://doi.org/10.1016/j.brainres.2005.12.015>
- Ito, T. A. (2011). Using ERPs to Understand the Process and Implications of Social Categorization. In J. Decety & J. T. Caciopo (Eds.), *The Oxford Handbook of Social Neuroscience* (pp. 445–522). New York: Oxford University Press.
- 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. <http://dx.doi.org/10.3758/CABN.5.1.21>
- Jones, E. E. (1990). *Interpersonal perception*. New York: W H Freeman/Times Books/Henry Holt & Co.

- Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology*, 52(3), 536–546. <https://doi.org/10.1037/0022-3514.52.3.536>
- Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G. H., & Miltner, W. H. R. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology*, 49(2), 207–219. <https://doi.org/10.1111/j.1469-8986.2011.01291.x>
- 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>
- Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The Price of Racial Bias: Intergroup Negotiations in the Ultimatum Game. *Psychological Science*, 24(12), 2498–2504. <https://doi.org/10.1177/0956797613496435>
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 1–33. <https://doi.org/10.18637/jss.v069.i01>
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique* (1st ed.). Cambridge, MA: The MIT Press.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>

- Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For Members Only: Ingroup Punishment of Fairness Norm Violations in the Ultimatum Game. *Social Psychological and Personality Science*, 5(6), 662–670.
<https://doi.org/10.1177/1948550614527115>
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510–532. <https://doi.org/10.1037/0033-2909.131.4.510>
- Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, 28(4), 441–448.
<https://doi.org/10.1016/j.neubiorev.2004.05.003>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
<https://doi.org/10.1037/0033-295X.84.3.231>
- Osinsky, R., Mussel, P., Öhrlein, L., & Hewig, J. (2014). A neural signature of the creation of social evaluation. *Social Cognitive and Affective Neuroscience*, 9(6), 731–736. <https://doi.org/10.1093/scan/nst051>
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25(3), 273–285.
<https://doi.org/10.3758/BF03211283>
- Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual*

Review of Sociology, 34(1), 181–209.

<https://doi.org/10.1146/annurev.soc.33.040406.131740>

Pager, D., Western, B., & Bonikowski, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, 74(5), 777–799.

<https://doi.org/10.1177/000312240907400505>

Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2010). Manipulation of feedback expectancy and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology*, 48(5), 656–664. <https://doi.org/10.1111/j.1469-8986.2010.01136.x>

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 118(10), 2128–2148.

<https://doi.org/10.1016/j.clinph.2007.04.019>

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, 52(4), 449–459.

<https://doi.org/10.1111/psyp.12370>

Qu, C., Wang, Y., & Huang, Y. (2013). Social exclusion modulates fairness consideration in the ultimatum game: an ERP study. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00505>

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.3). Vienna, Austria: R Foundation for Statistical Computing.

- Roese, N. J., & Sherman, J. W. (2007). Expectancy. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 91–115). New York, NY: Guilford Press.
- Rubin, Z., & Moore, J. C. (1971). Assessment of subjects' suspicions. *Journal of Personality and Social Psychology*, 17(2), 163–170.
<https://doi.org/10.1037/h0030299>
- Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., & Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *NeuroReport*, 16(4), 407.
- Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, 96(1), 58–83. <https://doi.org/10.1037/0033-295X.96.1.58>
- Stanford Center for the Study of Poverty and Inequality. (2017). State of the Union. *Pathways Magazine, Special Issue*. Retrieved from
<https://inequality.stanford.edu/publications/pathway/state-union-2017>
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, 111(1), 42–61. <https://doi.org/10.1037/0033-2909.111.1.42>
- Stanley, D. A., Phelps, E., & Banaji, M. (2008). The Neural Basis of Implicit Attitudes. *Current Directions in Psychological Science*, 17(2), 164–170.
<https://doi.org/10.1111/j.1467-8721.2008.00568.x>
- Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions.

- Proceedings of the National Academy of Sciences of the United States of America*, 108(19), 7710–7715. <https://doi.org/10.1073/pnas.1014345108>
- Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1589), 744–753. <https://doi.org/10.1098/rstb.2011.0300>
- Tortosa, M. I., Lupiáñez, J., & Ruz, M. (2013). Race, emotion and trust: an ERP study. *Brain Research*, 1494, 44–55. <https://doi.org/10.1016/j.brainres.2012.11.037>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Willadsen-Jensen, E. C., & Ito, T. A. (2008). A Foot in Both Worlds: Asian Americans' Perceptions of Asian, White, and Racially Ambiguous Faces. *Group Processes & Intergroup Relations*, 11(2), 182–200. <https://doi.org/10.1177/1368430207088037>
- Williams, D. R., & Mohammed, S. A. (2013). Racism and Health I: Pathways and Scientific Evidence. *American Behavioral Scientist*, 57(8), 1152–1173. <https://doi.org/10.1177/0002764213487340>
- Wu, Yan, & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research*, 1286(Supplement C), 114–122. <https://doi.org/10.1016/j.brainres.2009.06.032>
- Wu, Yin, Leliveld, M. C., & Zhou, X. (2011). Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. *Biological Psychology*, 88(2–3), 253–262. <https://doi.org/10.1016/j.biopsycho.2011.08.009>

- Wu, Yin, Zhou, Y., van Dijk, E., Leliveld, M. C., & Zhou, X. (2011). Social Comparison Affects Brain Responses to Fairness in Asset Division: An ERP Study with the Ultimatum Game. *Frontiers in Human Neuroscience*, 5.
<https://doi.org/10.3389/fnhum.2011.00131>
- Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP Correlates of Feedback and Reward Processing in the Presence and Absence of Response Choice. *Cerebral Cortex*, 15(5), 535–544. <https://doi.org/10.1093/cercor/bhh153>
- Yeung, N., & Sanfey, A. G. (2004). Independent Coding of Reward Magnitude and Valence in the Human Brain. *Journal of Neuroscience*, 24(28), 6258–6264.
<https://doi.org/10.1523/JNEUROSCI.4537-03.2004>
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151–175. <https://doi.org/10.1037/0003-066X.35.2.151>

Figures

Figure 1.

Figure 1. *Trust Game paradigm trial structure*. Every trial was separated by an inter-trial interval, during which a fixation cross was displayed. Participants played the role of the investor (referred to as Player A). At the beginning of each trial, participants were first shown a picture of the partner they would be playing against for that round/trial (2 sec). Then participants were allowed 2 sec to respond (via a button box) with how much of their allotted \$5 they would like to share with their partner for that particular round. Participants received positive feedback on 50% of trials within each Player B (trustee)

race condition (Black, White, Other). When participants failed to respond within the 2-second window, the feedback screen displayed: “Too slow!” ITI = intertrial interval.

Figure 2.

Figure 2. Heat maps identifying the location of the most positive peaks at the time point of each ERP peak signal. The yellow box highlights the locations of electrodes that were included in analyses for each ERP component. A note about RewP: Because RewP is a negative going positive slope, it is not the most positive peak at the time point of its maximal peak.

Figure 3.

Figure 3. Offer amount as a function of partner race and ATB scores. The y-axis depicts the average offer amount made during the trust game. Zero on the x-axis is the mean ATB score for the current sample, with higher scores indicating more negative attitudes toward Black people (i.e., more bias). The lines represent the predicted offer amount values as a function of ATB scores for Black (purple) and White partner trials (green). Shading represents the estimated 95% confidence bands. The scatter plot depicts average amounts offered to Black and White partners for each participant (i.e., two dots per person); it was necessary to collapse across trials in order to facilitate the legibility of the graph. The dashed grey vertical lines mark 1 *SD* above and below the mean ATB score

(i.e., the levels at which the simple slope effects were tested). Note: * denotes the simple slope was significant at $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$. ATB = Attitude Towards Blacks.

Figure 4.

Figure 4. Grand average ERP waveforms elicited by feedback (panels A and B) and faces (panel C) during the trust game. ERP waveforms were averaged across 15 frontal midline electrode sites for RewP (A), and averaged across 15 central parietal midline sites for P3 (B) and P2 (C). The highlighted grey region is the time window from which average amplitudes were quantified for each component. The x-axis is time in ms from the onset of the feedback stimulus (A and B) or face stimulus (C). Note that the y-axis is reverse scaled (consistent with ERP conventions). Purple lines represent Black partner trials, and green lines represent White partner trials. For panels A and B, solid lines denote feedback

indicating the partner kept the money and dashed lines denote feedback indicating the partner shared the money back with the participant.

Figure 5.

Figure 5. RewP amplitude as a function of partner race and offer amount. Zero on the x-axis represents the average amount offered by each participant. Separate lines represent the predicted RewP amplitude (y-axis) as a function of offer amount for Black partner trials (purple) and White partner trials (green), with 95% confidence bands (shaded region). The scatter plot depicts RewP averaged over trials and electrodes within each Offer Amount x Race cell per participant (i.e., up to 10 dots per participant). The dashed grey vertical lines mark 1 *SD* above and below the mean offer amount (i.e., the levels at which the simple slope effects were tested). Note: * denotes the simple slope was significant at $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$.

Figure 6.

Figure 6. RewP (left panel) and P3 (right panel) amplitude values as a function of partner race and attitudes toward Blacks. Zero on the x-axis is the mean ATB score for the current sample, with higher scores corresponding to more negative attitudes toward Black people (i.e., more bias). Lines representing the predicted amplitude values as a function of ATB scores are graphed separately for Black (purple) and White partner trials (green). The shaded region represents the 95% confidence bands. The scatter plot depicts average amplitudes (collapsed across trials and electrodes) for Black and White partner trials for each participant (i.e., two dots per person). Dashed grey vertical lines mark 1 *SD* above and below the mean ATB score. Note: * denotes the simple slope was significant at $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$. ATB = Attitude Towards Blacks.

Figure 7.

Figure 7. P2 amplitude elicited by faces as a function of partner race and attitudes toward Blacks. Zero on the x-axis is the mean ATB score for the current sample, with higher scores indicating more negative attitudes toward Black people (i.e., more bias). Lines that represent predicted P2 amplitude as a function of ATB scores are graphed separately for Black (purple) and White partner trials (green). The shaded region represents the 95% confidence bands. The scatter plot depicts average amplitudes (collapsed across trials and electrodes) for Black and White partner trials for each participant (i.e., two dots per person). Dashed grey vertical lines mark 1 *SD* above and below the mean ATB score.

Note: * denotes the simple slope was significant at $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$. ATB = Attitude Towards Blacks.

Figure 8.

Figure 8. Offer amount as a function of partner race and the difference in P2 amplitude elicited by Black and White faces (marginally significant interaction). Positive values on the x-axis signify a larger P2 for Black than White faces; negative values signify a larger P2 for White than Black faces. The solid lines represent predicted offer amounts for Black (purple) and White (green) partners as a function of individual differences in P2 race difference. The shaded region depicts the 95% confidence bands. The dots represent average offers made to Black and White partners for each participant (i.e., two dots per person) as a function of P2 race difference score. Dashed grey vertical lines mark 1 *SD* (1.70 μ V) above and below the mean P2 race difference score (1.17 μ V). Note: ‡

indicates the p -value for simple slope test at that predictor level was $< .1$, * indicates it was significant at $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. P2 Diff. = P2 race difference.

Appendix A

Stimuli Database Project Instructions

Who are the other players?

Now we are going to talk about who the other players are.

Thanks to a huge, NSF-funded, collaborative project, we have access to a rich database of pictures of real people that are linked to the actual responses made by those people.

[Tangent: Say this in your own words: (Don't look like you're reading)]

[You may have actually heard of this before - It's called the Dynamic Stimuli Database Project. It has been getting a decent amount of press lately – like, I think it was written up in the New York Times a few months ago. Anyways, ...]

These people participated previously as part of a massive, ongoing, real-response stimulus database-building project that has been taking place jointly across multiple labs at MU, Lincoln University – have you heard of Lincoln University?

[if they say “no,” say: It's a historically Black college in Jefferson City about 30 minutes south of Columbia.]

... and Washington University – you've heard of Wash U?

[if they say “no,” say: It's a large university located in St Louis.]

... and St Louis University (which is [obviously] in St Louis).

You'll be playing with some of these individuals today. They were part of a community population sample of over 1,000 people collected over the course of the last decade from Columbia, Jefferson City, St Louis, and the surrounding areas.

While the majority of psychology research is generally conducted using undergraduates, a great deal of effort was put forth in this project to collect a very diverse sample that was much more representative of the community at large. So we are very lucky to have access to such a rich database.

These individuals previously consented to the use of their photographs and response data for use in research investigating social decision-making processes. When they participated, they played both roles of Player A and Player B. As Player B, they were asked to respond to each possible offer that Player A could make. All of their responses have been stored in a database.

So when you make an offer as Player A during the game, the computer game program will give them feedback based on the real response made by that person, which may depend on the particular amount you decide to share with them.

The motivation for developing this database was to provide researchers with a more real-life social economic interaction that can be easily created and studied in the lab.

This database of realistic, real-life responses tied to the individuals who made those responses. This database is and has been used by researchers all over the world since the first stimuli set was developed in 2009.

So do you have any questions so far? About the database or the game you'll be playing? (Also, you'll have a chance to practice before we get started.)

Database:

As I already mentioned, this database project is currently ongoing. Before I apply the EEG cap, I wanted to actually talk to *you* about potentially taking part in this project. If you agree to participate in this project, you will be contributing to the advancement/improvement of the scientific methodology in this growing, new area of research.

Right now, we are specifically interested in recruiting college-age participants (since the current database consists of individuals from the community at large). Because in this research study, we are investigating the neural responses of Player A during the game, we are only interested in collecting your responses in the role of Player B for the purpose of the database. However, the development of such a database would be very valuable, because it would allow researchers to design multi-player studies (like the one you are presently taking part in) focusing on economic interactions among peers (since many labs use undergraduates in their studies).

If you agree to become part of the database, your answers will be completely confidential, and will be stored in the database in such a way that it will not be associated with any personal identifiable information (i.e., that could be linked/traced back to their identity). Just to be completely clear, only your answers during the first part of the study (as Player B) will be added to the mass database; the rest of the tasks you will be asked to complete today are for purposes confined to this laboratory, and that all responses are completely confidential.

Waiver/Consent:

I am now going to allow you to read over a separate waiver and consent form. Remember that participating in the database project is NOT a requirement for participating in this study. Please take some time to read over this consent form and fill it out. After you have read it, there is a box at the bottom of the form where you can indicate whether or not you'd like to take part in the database project.

If you decide not to take part in the database development project, we will still ask to take your picture and ask that you complete the task nonetheless, in order to ensure that all participants have a similar experience prior to beginning the primary portion of the experiment.

Appendix B

Dynamic Stimuli Database Project Waiver and Consent Form for Peer Database

What is the Dynamic Stimuli Database Project (DSDP)?

In research investigating economic decisions that involve social interactions, it can be challenging to efficiently recreate real-life interactions in the lab. The Dynamic Stimuli Database Project (or DSDP) represents a new and innovative strategy for addressing this problem. DSDP is an ongoing project to build and develop a large database of potential game interactions during the three most commonly utilized economic games in research across psychology and behavioral economics, such as the one you'll be playing today. The database consists of pictures of over 1,000 individuals in nearby areas across the Midwestern region of the United States. Like you, these individuals have previously participated in one of these three games, and their responses during one of these games are linked to their picture. This way, we have a database of realistic, real-life responses tied to the individuals who made those responses. This database is and has been used by researchers all over the world since the first stimuli set was developed in 2009.

Who is in charge of DSDP?

DSDP is a collaborative effort undertaken jointly by Washington University, St Louis University, University of Missouri-Columbia, and Lincoln University.

What is the purpose of the database for which I am being recruited?

The existing DSDP database contains pictures and responses from members of the community at large. However, we are currently developing a new database of individuals recruited from college campuses. Because so much research is conducted using undergraduate populations, such a database will be helpful for researchers specifically interested in multi-player economic games involving peers.

What will I be asked to do if I decide I would like to participate?

Your picture will be taken using a digital camera. You will subsequently be asked to complete a very brief computerized task, during which you will play the role of one player during a multi-player game.

What will happen to my picture and my responses?

Your answers will be completely confidential, and will be stored in the database in such a way that it will not be associated with any personal identifiable information (i.e., that could be traced back to their identity). Your responses along with your picture will only be made available to researchers, and will never be used for non-research purposes.

Signature:

Once you have read this consent form and your questions have been answered, please check one of the following. Would you like to contribute to the database? (Check one.)

- ☐ Yes, I consent to use of my photograph and my data in the Dynamic Stimuli Database Project database that will be distributed to other researchers.
- ☐ No, I DO NOT consent to use of my photograph and my data in the Dynamic Stimuli Database Project database that will be distributed to other researchers.

(Signature of participant)

(Date)

(Printed name of participant)

If you have agreed to contribute to the DSDP database, please complete the following form on the next page. Otherwise, please alert the experimenter that you are finished.

Photograph Release Form

I hereby authorize the Dynamic Stimuli Database Project, hereafter referred to as “DSDP,” to use the

photograph taken of me on _____ for use in research.
Date

I further acknowledge that my participation is voluntary and that I will not receive financial compensation of any type associated with the taking or use of these photographs or participation in research related activities. I acknowledge and agree that use of said photos confers no rights of ownership or royalties whatsoever.

I hereby release DSDP, its contractors, its employees, its associates, and any third parties involved in the creation and use of DSDP database stimuli sets, from liability for any claims by me or any third party in connection with my participation.

Authorization

(Signature of participant)

(Date)

(Printed name of participant)

Appendix C

Post-Experiment Questions

1. What do you think this study was about?

2. Did you use any strategies or “rules of thumb” to help you make decisions about how much money to share during the Trust Game? *If so, please explain.*

3. Even though Player B was different during each trial, did you feel like your responses were affected by the outcomes on previous trials (like, whether other players shared or kept the money)? Or were you able to judge each person individually?

4. Do you feel like you were pretty good at judging character during the game? Or were you often surprised by how Player B responded?

5. How much did you care about earning money during this experiment? Circle your answer on the scale below.

1	2	3	4	5
Not at all				Very much

6. How much did it bother you when Player B kept all of the money for himself? Circle your answer on the scale below.

1	2	3	4	5	6	7
Not bothered at all			Moderately bothered			Very bothered

7. How boring did you find this experiment to be? Circle your answer on the scale below.

1	2	3	4	5	6	7
Not boring at all			Moderately boring			Extremely boring

8. Did the game feel like it went on for too long, so that it made it difficult for you to continue to pay attention towards the end?

9. Did you feel like you were still able to pay attention towards the end of the experiment?

10. Do you have any additional comments that you would like to share with the experimenter?

Appendix D

Funneled Debriefing Discussion/Interview Questions

The following questions are a general outline for debriefing and probing for suspicion. The actual questions asked to each individual may differ depending on how the individual participant responds to each question. These questions are designed to get the participant to speak about their experience participating in the experiment without asking leading questions, and so the conversation goes in the direction that the participants choose. Start by reading the Post-Experiment Questionnaire responses before you begin this process. (Note that if their verbal answers just reiterate things they already wrote in the post-experiment questionnaire, you don't have to write it down.)

1. **Did you have enough time to respond during the task?** YES NO
2. **Did you feel like the task was really long?** YES (too long) NO
3. **Were you still able to pay attention at the end?** YES NO
4. **Have you heard about this study from anyone in advance?** YES NO
5. **Have you played an economics game like the Trust Game before?** YES
NO
6. **Were you often surprised a lot of the time at what people actually did? Or do you feel like you were pretty good at this game?**

Surprised a lot Did pretty well Both Other:

7. **What do you think the main purpose of the study was?** *(Note: Look at response to Question #1 on the Post-Exp. Questionnaire – if P says anything about stereotypes or race, ask them about that: e.g., What made you think that the study was about ____ (stereotypes, bias, etc)?.)*

Related to race NOT related to race

(If related to race:) What made you think that?

8. **Did any of the tasks seem odd and make you question why you were being asked to do them?**

YES NO

If so, which?

If a task was odd, what do you think the purpose of the task was?

9. Did you actually believe there was a possibility that you could receive money at the end of the experiment? *[If “NO,” inquire further to ensure that they don’t just think they performed badly during the task, and that they actually believed we were lying to them about the money.]*

YES

NO

10. If you had to guess, what percent of time would you say that the other players were sharing back with you?

11. Did you feel like you were able to treat each person individually, and remember that previous trials had no effect on the current trial?

YES

NO

12. Did you feel like you were making decisions with real people? Or did you feel like you were playing against the computer?

Real people

Computer

Debriefing

Database Deception:

Actually, you were in fact playing against the computer – the pictures were not tied to responses at all. Instead the feedback you received was randomly generated by the computer, in such a way that you received positive feedback 50% of the time. A database such as the one that I described unfortunately does not exist. As such, we actually are not going to add your picture to a database – it actually will not be used for anything. We apologize for deceiving you.

The reason why we found it necessary to deceive you is that we are trying to model real-world behavior and social interactions. You can imagine that how you might behave while making decisions about investing money with fake people (or a computer) would probably be very different from how you might act while making decisions involving real people. As such, it was very important for the purpose of this experiment that you believed that you were playing with real people. We believe that there is no other way that we could have effectively tested our hypotheses without deceiving you in some way, and we hope you understand our reasoning for doing so.

About the Study:

The purpose of this experiment is to examine how unconscious, automatic racial bias contributes to economic decisions that depend on trustworthiness.

We are expecting to find that during the Trust Game, participants are more trusting of (meaning that they share more money with) White players than Black players. However, we don't believe that participants will be doing this intentionally. Instead, we think that this behavior is simply the manifestation of unconscious, subtle biases against Black people, or even the awareness of negative stereotypes about Black men.

We are also interested in a particular brainwave that is thought to track what is known in economics as a “prediction error.” This term refers to the difference between what you thought was going to happen, and what actually happens. So when there is a big difference between what you were expecting the other player to do, and what they actually did, this would produce a much larger response in this particular brainwave. We are expecting to find that when Black players reciprocate generosity, there will be a much larger brain response compared to when White players reciprocate, suggesting that participants are more surprised in response to Black players reciprocating.

Do you have any questions?

REMEMBER: Request that they not discuss this with anyone in their class: It is important that people not know what this study is about (or that it pays money) ahead of time.

Supplemental Methods

Scoring the Global Suspicion Variable

For all variables, 0 was assigned when participants did not express suspicion in the ways defined below.

Suspicion of other players. During the funneled debriefing, participants were asked whether they felt like they were playing against a computer instead of real people. If they answered yes, they were further questioned about the extent to which they believed the DBSP manipulation (i.e., that the feedback they received during the trust game was tied to real behavioral responses given by the people depicted in the pictures). Fourteen individuals reported feeling like they were playing against a computer, but that they still believed the DBSP manipulation – these responses were coded as 1. Five participants who expressed doubts about whether the DBSP backstory was true at all were coded as 2.

Suspicion of receiving money. During debriefing, the experimenter asked participants, “Did you actually believe there was a possibility that you could receive money at the end of the experiment?” Participants who expressed uncertainty ($n = 6$) were assigned a score of 1. Participants who admitted they did not believe the experimenter (or believed the experimenter was mistaken) and thought there was no possibility of receiving money ($n = 6$) were assigned a score of 2.

Suspicion of race hypothesis. During the first post-experiment questionnaire participants completed following completion of the trust game, they were asked, “What do you think the study was about?” I used this open response question to assess expressed awareness of the race hypothesis. Importantly, they answered this question before having

seen our three questionnaires about racial bias. Participants who stated they thought the experiment was about either a) race listed among other features (such as age, facial features, etc), or b) stereotypes with no mention of race specifically ($n = 15$), were coded as 1. Responses that mentioned race (or ethnicity, etc.) in isolation ($n = 10$) were coded as 2, because it suggested these participants might have been more focused on race in particular (as opposed to just listing race as one aspect of physical appearance).

Supplemental Table S1. Random effects in multilevel models.

DV	Model	Random Effects			
		Subject	Trial	Face Picture	Electrode
<i>Offer Amount</i>	<i>Race</i>	Intercept, Slope: Race	-	Intercept, Slope: Race	N/A
	<i>Race (White participants only)</i>	Intercept, Slope: Race	-	Intercept	N/A
	<i>Race x ATB</i>	Intercept, Slope: Race	-	Intercept	N/A
	<i>Race x Suspicion</i>	Intercept, Slope: Race	-	Intercept	N/A
	<i>(Race x P2 Race Difference) + (Race x P2 Within Subjects)</i>	Intercept, Slope: Race	-	Intercept	N/A
<i>RewP</i>	<i>Race x Feedback</i>	Intercept, Slopes: Race x Feedback	Intercept, Slopes: Race, Feedback	Intercept	Intercept
	<i>Race x Feedback x Offer (within-subjects)</i>	Intercept, Slopes: Race x Feedback x Offer	Intercept	Intercept	Intercept
	<i>Race x Feedback x ATB</i>	Intercept, Slopes: Race x Feedback	Intercept	Intercept	Intercept
<i>P3</i>	<i>Race x Feedback</i>	Intercept, Slopes: Race, Feedback	Intercept, Slopes: Race, Feedback	Intercept, Slopes: Race, Feedback	-
	<i>Race x Feedback x Offer (within-subjects)</i>	Intercept, Slopes: Race x Feedback x Offer	Intercept	Intercept	Intercept
	<i>Race x Feedback x ATB</i>	Intercept, Slopes: Race x Feedback	Intercept, Slope: Trial	Intercept	Intercept
<i>P2</i>	<i>Race</i>	Intercept, Slope: Race	Intercept	Intercept	Intercept
	<i>Race x ATB</i>	Intercept, Slope: Race	Intercept	Intercept	Intercept

Note. When fixed effect predictors or random effect slopes include interactions, all affiliated lower order terms were also included. A dash indicates that the intercept of the

random effect was not significant, and so was not included in the model. DV = dependent variable; N/A = Not applicable.

Vita

Meredith P. Levsen (Johnson) was born in New York, NY, and raised in Durham, NC. She graduated from high school at Camelot Academy in Durham, NC in 2004. She then attended New York University (NYU) and graduated in 2008 with a Bachelor of Arts degree in Psychology with a minor in Africana Studies. She worked as a Lab Manager/Research Assistant at Rutgers University – Newark in Newark, NJ for three years after graduation. Meredith began her graduate work in Psychology at the University of Missouri in August 2012. She earned her Doctor of Philosophy in Social & Personality Psychology with a graduate minor in Psychological Statistics and Methods in May 2018.