

Dynamic neural mechanisms underlie race disparities in social cognition



Brittany S. Cassidy*, Anne C. Krendl

Indiana University, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:

Received 21 August 2015

Accepted 13 February 2016

Available online 22 February 2016

Keywords:

Race perception

Trustworthiness

Face processing

Differentiation

ABSTRACT

Race disparities in behavior may emerge in several ways, some of which may be independent of implicit bias. To mitigate the pernicious effects of different race disparities for racial minorities, we must understand whether they are rooted in perceptual, affective, or cognitive processing with regard to race perception. We used fMRI to disentangle dynamic neural mechanisms predictive of two separable race disparities that can be obtained from a trustworthiness ratings task. Increased coupling between regions involved in perceptual and affective processing when viewing Black versus White faces predicted less later racial trust disparity, which was related to implicit bias. In contrast, increased functional coupling between regions involved in controlled processing predicted less later disparity in the differentiation of Black versus White faces with regard to perceived trust, which was unrelated to bias. These findings reveal that distinct neural signatures underlie separable race disparities in social cognition that may or may not be related to implicit bias.

© 2016 Elsevier Inc. All rights reserved.

Race disparities are wide-ranging in behavior, from less perceived trust in Black versus White faces (Stanley et al., 2011), lower ratings on job performance for Black versus White managers (Greenhaus et al., 1990), and broad deficits in recognizing other- versus same-race faces (Meissner and Brigham, 2001). Although cultural associations of Black males with aggression (Dovidio et al., 1996) may potentially underlie these tensions by perpetuating implicit bias toward minorities (e.g., McConnell and Leibold, 2001), other work suggests that race disparities emerge in several ways, some of which may be relatively independent of implicit bias. Disentangling the mechanisms underlying race disparities that are and are not related to implicit bias is critical to develop strategies that most effectively mitigate their consequences.

A wealth of social psychological research has demonstrated the pernicious effects of implicit bias on behavior (Gawronski et al., 2003). For instance, individuals with higher versus lower levels of implicit bias (as measured through subtle measures of prejudice) evaluate Black faces as less trustworthy than White (Stanley et al., 2011), discriminate more against Black proposers in economic games (Kubota et al., 2013; Stanley et al., 2011), have tenser intergroup interactions (McConnell and Leibold, 2001), and have more stereotyped mental representations of outgroup members (Dotsch et al., 2008). In contrast, implicit bias does not predict own-race effects in memory for faces (e.g., Slone et al., 2000), while the ability to distinguish (i.e., differentiate) other-race faces from one another does predict these effects (Goldstein and Chance, 1985; Hills and Lewis, 2006). For

example, given a majority of Black professional basketball players, White fans better recognize Black faces than do White novices because they can better differentiate between them (Li et al., 1998). Even though differentiation itself may be separable from bias (Ferguson et al., 2001), less differentiation of outgroup versus ingroup members has downstream consequences on behavior. Indeed, individuals who do not differentiate other-race faces as well as others do not recognize them as well as ingroup members (Goldstein and Chance, 1985) and stereotype them more (Linville and Fischer, 1998).

However, it is unclear whether distinct components of race perception underlie race disparities due to bias or differentiation. To distinguish mechanisms for these disparities, our study dissociated bias-related disparities from disparities relatively independent from bias based on performance in a single task. The benefit to using the same task to disentangle neural mechanisms for these race disparities is that it controls for differences due to task. When White individuals view Black and White faces, they engage multiple processes reflected in neural activity (Amodio, 2014). For instance, they differentiate faces (Linville et al., 1989) as well as evaluate them (Stanley et al., 2012). In the case of the former, these processes emerge from perceiving race, but are not necessarily influenced by bias (Meissner and Brigham, 2001), whereas bias does influence evaluative judgments (Stanley et al., 2011). Because neuroimaging has importantly localized brain regions contributing to perceptual, affective, and cognitive components of race perception and how implicit bias impacts those regions' engagement (for reviews, see Amodio, 2014; Kubota et al., 2012), it allows us to disentangle how neural activity contributes to later behavioral race disparities. This is a critical consideration because understanding how different race disparities arise provides a basis to develop interventions that reduce them.

* Corresponding author at: Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St., Bloomington, IN 47405, USA.
E-mail address: bscassid@indiana.edu (B.S. Cassidy).

Perceiving race engages affective processing regions, including the amygdala and orbitofrontal cortex (OFC). Imaging work on implicit bias has also found amygdala and OFC activity correspondence to ingroup favoritism (Beer et al., 2008), suggesting these regions' responses to race may underlie individual differences in bias-related race disparities. The amygdala has been implicated widely in emotional face processing (Vuilleumier et al., 2001), with activity toward Black versus White individuals reflecting threat (Chekroud et al., 2014). Amygdala response to Black over White faces increases with more implicit bias (Phelps et al., 2000), suggesting greater responses to Black versus White faces may underlie exacerbation of later bias-related race disparities.

OFC activity also contributes to affective face processing (Vuilleumier et al., 2001), particularly in cues that guide approachability judgments related to trust (Willis et al., 2010). OFC activation reflects affect-based judgments with regard to race in that it decodes representations of Black and White faces during friendship evaluation (Gilbert et al., 2012), and potentially may represent racial groups during evaluation better than the amygdala (Amodio, 2014). Although less studied in race perception than other regions, research suggests a strong role of OFC in group preferences that reflect the representation of value (Kringelbach and Rolls, 2004) such that stronger OFC responses to ingroup versus outgroup members predict biases in liking of these individuals (Van Bavel et al., 2008). This positive outcome from stronger OFC response allows for the possibility that stronger OFC response to Black versus White faces may underlie attenuation of bias-related evaluative race disparities later.

Implicit bias also alters perceptual processing of other- versus same-race faces. Most studies on this topic have focused on the fusiform gyrus, a region widely implicated in face processing (Kanwisher et al., 1997), and for which viewing own- versus other-race faces enhances its response (Golby et al., 2001). Critically, White individuals with higher levels of bias have increased fusiform activity in response to White as compared to Black faces, whereas lower bias individuals do not exhibit a difference (Brosch et al., 2012). However, the fusiform gyrus has also been implicated in non-bias-related behaviors associated with race. For instance, activity in this region contributes to own-race memory biases driven by differentiation, such that lower activity toward outgroup versus ingroup members may reflect perceptual homogeneity and failures to encode outgroup members at an individual level (Golby et al., 2001). These findings suggest that enhanced fusiform activity to Black over White faces may underlie the later attenuation of bias-related and bias-independent race disparities. However, even though the fusiform may engage regardless of implicit bias, the connectivity of the fusiform to different regions involved in affective or cognitive processing may differ in predicting race disparities related to bias versus not.

Simple race perception tasks also elicit activity from regions implicated in cognitive control, and specifically lateral prefrontal cortex (Amodio, 2014). Although many past neuroimaging tasks were not designed to assess control (e.g., Cunningham et al., 2004), past work has identified a link between dorsolateral prefrontal cortex (dlPFC) activity and controlling race-related stereotypes (Amodio, 2010). Prior work shows that differentiation requires motivation and cognitive effort (Hills and Lewis, 2006), suggesting that increased dlPFC activity toward Black versus White faces may underlie later attenuation of race disparities in differentiating faces that are relatively independent from bias. Indeed, individuals engage prefrontal regions when processing faces they are more likely to differentiate (Feng et al., 2011). Prefrontal engagement has also been implicated in regulating prejudiced responses (Amodio, 2014). For instance, control regions are more active in individuals with high versus low implicit bias when they evaluate Black versus White faces (Richeson et al., 2003). More activation among higher bias individuals suggests they may need more control to inhibit their bias (Devine, 1989) and comply with egalitarian social norms (Richeson et al., 2003).

Overall, these findings suggest that bias impacts regions involved in affective processing during race perception. While bias may also influence activity in perceptual and cognitive processing regions during race perception, perceptual and cognitive associations with bias may be influenced by connectivity to affective processing regions while perceiving race. These findings also suggest that the contributions of these components of race perception to later race disparities may depend on the disparity being assessed. To this end, we disentangled neural mechanisms predictive of different race disparities by connecting neural activity in a simple race perception task to two separable aspects of a later trustworthiness ratings task: evaluative racial trust disparity (i.e., perceiving less trust in Black versus White faces) and racial differentiation disparity (i.e., differentiating Black less than White faces). Differing from paradigms tracking activity during online evaluations, we assessed whether variations in neural response while perceiving race support later behavioral disparities within the same individuals. We assessed trustworthiness because perceivers automatically extract information from faces regarding their trustworthiness (e.g., Meconi et al., 2014; Todorov, 2008), suggesting reliability between brain activity when perceiving faces and later evaluations. These evaluations occur in the absence of cognitive control (e.g., Rule et al., 2013; Todorov, 2008), drawing on perceptual and affective processes (e.g., Winston et al., 2002). Connecting to race biases, trustworthiness evaluations draw from approach responses (Todorov, 2008) much like evaluative associations with race that are distinct from stereotype-based biases (Amodio and Devine, 2006). At the same time, people make distinctions between faces by way of their scale use when making a series of trustworthiness evaluations (Linville et al., 1989). Differentiating faces requires motivation and cognitive effort (Hills and Lewis, 2006), suggesting differentiation draws more on control than automatic evaluations.

In sum, we predicted that more fusiform and OFC activity toward Black versus White faces would correspond with less racial trust disparity in later evaluations. Further, we expected that greater connectivity between these regions would correspond with less racial trust disparity, reflecting bias-related behavior as dynamically driven by perceptual and affective processes. In contrast, given the association between amygdala response to Black versus White faces and threat (Chekroud et al., 2014), we expected greater amygdala activity toward Black versus White faces to underlie more racial trust disparity later. In contrast, we anticipated more fusiform or lateral prefrontal activity toward Black versus White faces to correspond with less later racial differentiation disparity given perceptual and cognitive contributions to differentiation. We also expected that more connectivity between regions reflecting these processes would correspond with less racial differentiation disparity.

Methods

Participants

Thirty right-handed White adults with no history of neurological problems (18–29 years, 17 female; $M_{\text{age}} = 21.27$, $SD = 2.38$) recruited from Indiana University participated and provided informed consent. This sample size was derived on the basis of past neuroimaging studies on race perception (for a review of these studies, see Amodio, 2014). The Indiana University IRB approved this study. Participation was completed over a pre-testing day and a separate scanning day.

Procedure

Pre-testing

Participants completed an extensive fMRI screening and measures relevant to the present task and others in the laboratory. The behavioral measures completed in the lab included the Implicit Association Test (IAT) for race (Greenwald and Banaji, 1995; Greenwald et al., 2003) –

which uses reaction time (RT) to quantify implicit race bias – the Attitudes Toward Blacks (ATB; Brigham, 1993) – which measures explicit attitudes toward Black individuals (with higher scores indicating increased explicit anti-Black bias) – and Internal (IMS) and External (EMS) Motivation to Respond Without Prejudice (Plant and Devine, 1998) – which assesses how motivated people are by internal and external, respectively, sources to not appear prejudiced (with higher scores indicating increased motivation).

For the IAT, participants viewed 20 male faces (10 Black and 10 White not used in the full experiment) and 20 words (10 pleasant and 10 unpleasant) and categorized them in a stereotypically congruent (e.g., a pleasant word paired with a White face) or incongruent (e.g., a pleasant word paired with a Black face) way. The IAT followed the seven-block protocol described in Greenwald et al., 2003. Congruent and incongruent blocks were presented in a pseudorandom order, meaning half of the participants saw the congruent block first and half saw the incongruent block first. After removing trials less than 200 ms and over 3000 ms, we used IAT D (Greenwald et al., 2003), which divides the RT difference between incongruent and congruent blocks by the standard deviation of all latencies, as a measure of implicit race bias. See Table 1a for descriptive statistics on these measures.

Scanning

Participants completed the fMRI study described below as well as a separate study that did not concern race perception. The order in which participants completed these studies was counterbalanced. Participants made age judgments in response to each face during the fMRI race task (in order to ensure they were attending to the stimuli) and were therefore unaware that the purpose of this study was to assess race-based differences in perceived trust. After scanning, participants rated all faces from the race fMRI study on their trustworthiness and stereotypicality in a self-paced task. Ratings were collected on a 7-point scale (e.g., 1 = not at all trustworthy to 7 = very trustworthy). The order of making trustworthiness and stereotypicality ratings was counterbalanced among participants.

fMRI stimuli and task

Ninety Black and 90 White young male faces with neutral expressions were drawn from the Eberhardt Face Database (<http://www.stanford.edu/group/mcslab/cgi-bin/wordpress/examine-the-research/>), a database used in prior racial trust disparity research (Stanley et al., 2011). Prior online ratings of 35 individuals who rated the Black faces and 33 who rated the White faces showed the selected Black and White faces did not differ in their attractiveness, $p = .76$. Prior online ratings of 19 individuals who rated the Black faces and 19 who rated the White faces were used to categorize faces as stereotypically low, medium, or high with respect to their race. Although stereotypicality could promote differentiation to an extent, the stereotypicality manipulation

was of interest for a separate study and not relevant to testing the present hypotheses. No interactive effects of stereotypicality with race occurred in the discussed regions. As such, stereotypicality will not be discussed in further depth here because it is beyond the scope of this paper. However, faces varying on stereotypicality were modeled separately in our first-level design matrix given that they were built into the study design.

The task was modeled as an event-related design over two runs lasting three minutes (90 TRs) each. Participants viewed images (45 Black and 45 White faces in each run, as well as 15 instances of a Black and White randomly checked visual noise image) for 1 s each. The noise image was included for the option of comparing face-related brain activity to non-face visual stimuli, but this comparison was not used in the present analyses. Images were randomly presented, and all conditions were equally represented in both runs. The order of stimuli and fixations were created using a random number generator and by manual inspection to verify that no two stimuli of the same type (e.g., low stereotypicality Black faces) appeared twice in succession and that no stimuli of the same race appeared more than three times in succession. For each trial, participants categorized whether they thought a face was younger or older than 24 years old during the 1-s presentation of the face. The goal of these ratings was to identify brain regions engaged when participants made categorical judgments non-specific to race about Black and White faces. Because in everyday life individuals do not typically make explicit race judgments on Black and White individuals (e.g., “is this person Black or White?”), the fMRI race literature typically uses such judgments during fMRI tasks (e.g., Cunningham et al., 2004; Hart et al., 2000; Richeson et al., 2003). Age categorization is one such task in which individuals evaluate individuals without explicitly evaluating their race (e.g., Ronquillo et al., 2007; Wheeler and Fiske, 2005).

Periods of jitter, in the form of a fixation cross at the center of the display, ranged from 1 to 7 s and were pseudorandomly presented throughout the task. More specifically, there were five 1-s fixations, six 3-s fixations, four 5-s fixations, and 2- to 7-s fixations in each run ($M_{\text{jitter}} = 3.35$ s; $SD = 2.03$), with 8 s of fixation at the beginning and 10 s at the end, for a total of 75 s of fixation and 105 s of stimulus presentation during the task. Four dummy scans were also included at the start of each run in order to allow for stabilization of the scanner signal. These were excluded from analyses.

Participant responses were monitored to ensure attention during the task. Suggesting sufficient attention to the task, participants responded to an average 171.63 ($SD = 8.32$) of 180 trials for an average response rate of 95.35% ($SD = 4.62\%$). Participants responded that a face was younger than 24 an average of 93.27 times ($SD = 21.08$), and 78.37 times ($SD = 20.45$) that a face was older than 24. Before scanning, participants completed ten 1-s practice trials during which they made age judgments of five Black and five White female faces. These faces were unique from those used during the scanner task.

Table 1
Behavioral data summary.

A. Mean (SD) values for dependent variables and questionnaire measures						
	Racial trust disparity	Racial differentiation disparity	IAT D	ATB	IMS	EMS
	.20 (.65)	.02 (.04)	.43 (.36)	43.52 (12.15)	38.28 (6.10)	29.66 (9.94)
B. Correlations (r) between dependent variables and questionnaire measures						
	Racial differentiation disparity	IAT D	ATB	IMS	EMS	
Racial Trust Disparity	.06	.39*	.57**	-.28	.19	
Racial Differentiation Disparity		-.001	-.29	.46*	.12	
IAT D			.29	-.25	.21	
ATB				-.72**	.27	
IMS					-.08	

* $p < .05$.

** $p < .01$.

Data acquisition and analysis

Whole-brain imaging was performed on a Siemens 3.0T TIM Trio MRI scanner at the Indiana University Imaging Research Facility in Bloomington, Indiana. Anatomical images were acquired with a high-resolution 3-D magnetization prepared rapid gradient echo sequence (224 slices, TE = 3.02 ms, TR = 2200 ms, flip angle = 9°, .8 × .8 × .8 mm voxels). Functional images were collected over two runs of 90 time points each, using a fast field echo-planar sequence sensitive to blood oxygen level-dependent contrast (T2*; 32 slices with 3.5 mm thickness and 3.5 mm skip, TR = 2000 ms, TE = 30 ms, flip angle = 70°).

Preprocessing and analyses of functional data were conducted in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Images were slice-time corrected, realigned to correct for motion, normalized to the MNI (Montreal Neurological Institute) template, and smoothed using a 8-mm FWHM isotropic Gaussian kernel. Data were resampled to 3 mm-isotropic voxels in a 96 × 96 matrix. Using custom artifact detection software to detect motion artifact (http://www.nitrc.org/projects/artifact_detect), runs were analyzed on a participant-by-participant basis to detect outlier time points. We excluded volumes during which participant head motion exceeded 1 mm and volumes in which the overall signal for that time point fell three standard deviations outside the mean global signal for the entire run. Outlier time points were excluded from the GLM analysis via the use of participant-specific regressors of no interest. A general linear model incorporating the seven image types and covariates of no interest (a session mean, a linear trend, six movement parameters derived from realignment corrections), computed parameter estimates (β), and t -contrast images (containing weighted parameter estimates) for each comparison at each voxel and for each participant.¹ In the GLM, event duration was modeled as 1 s to account for the entire presentation of each stimulus.

Whole-brain analyses

In order to determine how brain activity to Black over White faces varied with race-related behavior, we created the contrast [Black > White] for each participant. First-level [Black > White] images were submitted to the second-level covariate analyses. Participant-level covariates of interest were racial trust disparity (indicated by post-task ratings; see below for calculation) and racial differentiation disparity (indicated by post-task ratings; see below for calculation). Consistent with the little prior work on the neural correlates of race differences in trust, peak coordinates were identified by an extent threshold of 15 contiguous voxels (resampled) exceeding a voxel-wise threshold of $p < .005$ (Stanley et al., 2012). One-thousand Monte Carlo simulations indicated this provided a corrected experiment-wise threshold of $p < .05$. This method (described in Slotnick et al., 2003) has been used in numerous social cognition studies. We identified the

¹ We analyzed RTs to age decisions in a 2 (Race: Black, White) × 2 (Age: younger, older) repeated-measures ANOVA. No main effects of Age or Race emerged, $ps > .91$. An unexpected Age × Race interaction emerged, $F(1, 29) = 9.29, p = .01, \eta_p^2 = .24$. Participants had longer RTs toward Black faces categorized as younger ($M = 683.84, SD = 71.39$) versus older ($M = 653.90, SD = 54.28$), $F(1, 29) = 4.41, p = .05, \eta_p^2 = .13$. In contrast, participants had longer RTs toward White faces categorized as older ($M = 683.91, SD = 62.90$) versus younger ($M = 655.65, SD = 45.63$), $F(1, 29) = 5.25, p = .03, \eta_p^2 = .15$. Critically, the differences in RTs for younger versus older faces for both Black and White faces did not correlate with racial trust disparity, racial differentiation disparity, IAT D, IMS, EMS, or ATB, $ps > .12$. To further verify that age categorization did not unduly influence trustworthiness evaluations by race, we tested for the presence of an Age (younger, older) × Race (Black, White) interaction on trustworthiness evaluations. No interaction emerged, $F < 1, p = .55$. These analyses suggest that the age categorization task did not unduly influence trustworthiness evaluations by race. Thus, consistent with prior race perception work using simple categorization tasks (e.g., Hart et al., 2000; Ronquillo et al., 2007; Wheeler and Fiske, 2005), we did not include age decisions or reaction times (RTs) to these decisions in our first-level model.

anatomical location and Brodmann area of each emergent region using the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002) within MRICron (Rorden and Brett, 2000). To justify and aid in our interpretation of emergent activated regions, we used NeuroSynth (www.neurosynth.org; Yarkoni et al., 2011) to verify that similar activations found in other experiments reflected the hypothesized perceptual, affective, and cognitive processes.

Psychophysiological interaction analyses

Psychophysiological interaction (PPI) analyses performed using gPPI (<http://www.nitrc.org/projects/gppi>) (McLaren et al., 2012), examined functional connectivity from seed regions identified by the whole-brain analyses (described below) also varying with individual differences in race-related behavior. The deconvolved time series from a 6 mm radius spheres around the coordinates of these seed regions were extracted from each participant. First-level images were entered into the second-level covariate analyses to identify regions whose functional connectivity to seed regions varied by race-related behavior when perceiving Black versus White faces. We again identified regions with an extent threshold of 15 contiguous voxels (resampled) exceeding a voxel-wise threshold of $p < .005$ for a corrected experiment-wise threshold of $p < .05$.

Results

Racial trust disparity

Behavior

Like prior work (Stanley et al., 2011), we standardized ratings for each participant around their mean rating. A paired-samples t -test revealed no overall difference in the perceived trustworthiness of White faces ($M = .10, SD = .33$) and Black ($M = -.10, SD = .33$) faces, $t(29) = 1.67, p = .11$, replicating prior work (Stanley et al., 2011). We subtracted the mean of standardized ratings of Black faces from the mean of standardized ratings of White faces to quantify relative racial trust disparity in each participant. Critically replicating prior work (Stanley et al., 2011), individuals with more implicit race bias expressed more racial trust disparity, $r(28) = .39, p = .03, 95\% \text{ CI } [.035 .66]$ (Fig. 1a). See Table 1b for correlations between dependent variables and questionnaires.

fMRI

We first assessed regions whose response to Black vs. White faces varied by individual differences in racial trust disparity in a covariate analysis (Table 2a). Consistent with our hypotheses, this analysis revealed right orbitofrontal cortex (OFC) activation (Fig. 1b). Relatively more activation toward Black versus White faces predicted later reduced racial trust disparity.

PPI revealed regions whose functional connectivity with the right OFC seed (peak MNI coordinates 18, 66, 0), identified by the above whole-brain analysis, also varied by individual differences in racial trust disparity when perceiving Black versus White faces. As predicted, functional coupling with the seed region varying by individual differences in racial trust disparity emerged in right fusiform gyrus (Fig. 1c, Table 2b). Simply put, more fusiform-OFC connectivity when perceiving Black versus White faces predicted less later racial trust disparity.

Racial differentiation disparity

Behavior

To assess the differentiation of perceived trustworthiness in Black and White faces, we calculated the probability of discriminating between category members separately for Black and White faces using the distribution of scale use in the trustworthiness ratings task. We calculated differentiation using the probability of differentiation (P_D),

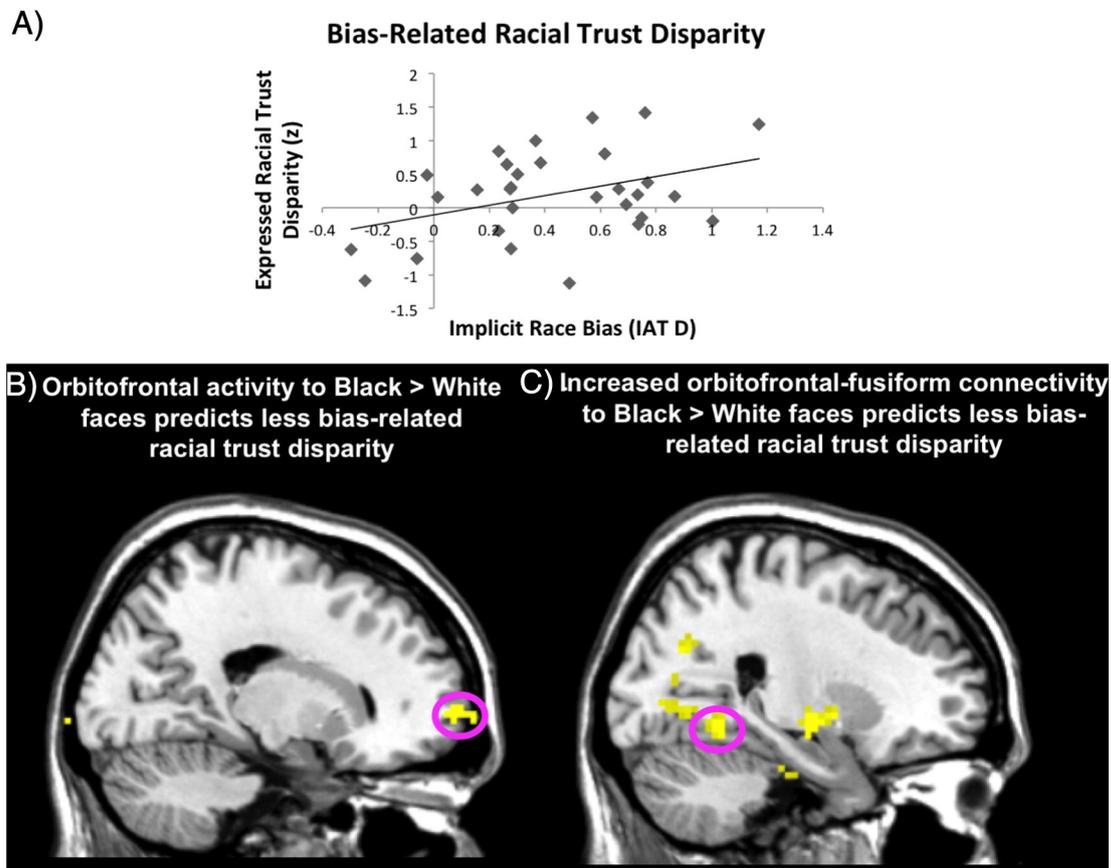


Fig. 1. The magnitude of bias-related racial trust disparity (A) depended on individual differences in orbitofrontal response to Black versus White faces (B) and orbitofrontal-fusiform connectivity when perceiving Black versus White faces (C).

where i is the level on the scale and P is the proportion of ratings at the n th level (Linville et al., 1989):

$$P_D = 1 - \sum P_i^2$$

Here, P_D for Black faces refers to the probability that two randomly chosen Black faces would differ in perceived trustworthiness, and P_D for White faces refers to the probability that two randomly chosen White faces would differ on this attribute. A paired-samples t -test revealed less differentiation of outgroup members ($M = .76$, $SD = .05$) versus ingroup members ($M = .78$, $SD = .05$), $t(29) = 2.30$, $p = .03$, $d = .42$. We subtracted the P_D of Black faces from the P_D of White faces for each participant to quantify racial differentiation disparity.

fMRI

We first assessed regions whose response to Black versus White faces varied by individual differences racial differentiation disparity (Table 3a). As hypothesized, activity in lateral prefrontal regions (Fig. 2a) varied by racial differentiation disparity. Individuals with more dorsolateral and ventrolateral prefrontal activity toward Black versus White faces had less racial differentiation disparity later.

We first focused our PPI analyses on the right dlPFC seed (peak MNI coordinates: 24, 42, 21) identified by the whole-brain analysis because its relationship with differentiation disparity was more robust than the other emergent right dlPFC region. PPI revealed regions whose functional connectivity with the seed while perceiving Black versus White faces varied by individual differences in racial differentiation disparity (Table 3b). Interestingly, enhanced connectivity to the seed when perceiving Black versus White faces that varied by racial differentiation

disparity emerged in left middle cingulate cortex (Fig. 2b). Simply put, more functional coupling between dlPFC and cingulate cortex when perceiving Black versus White faces predicted less disparity in differentiating them later. PPI analyses from the left ventrolateral prefrontal seed (peak MNI coordinates: $-30, 51, 3$) and a second dorsolateral seed (peak MNI coordinates: $27, 48, 15$) yielded no significant activations.

Discussion

Distinct and separable patterns of neural activity when perceiving race predicted later obtained racial trust disparity related to implicit bias and racial differentiation disparity unrelated to bias. These relationships occurred even without participant awareness that they would later rate faces on their trustworthiness. Specifically, when perceiving race, the engagement of brain regions shown by related work to reflect perceptual and affective components of race perception (Amodio, 2014) predicted later racial trust disparity, a behavior related to implicit bias (Stanley et al., 2011). Moreover, the engagement of brain regions previously shown to reflect a control-related component of race perception (Amodio, 2010) predicted later racial differentiation disparity.

Our finding that higher implicit bias predicts more expressed racial trust disparity replicated prior work (Stanley et al., 2011). We extended this work by demonstrating that the OFC response to Black over White faces predicted less racial trust disparity later. In the context of past imaging work on OFC and work showing effects of emotion on intergroup attitudes (DeSteno et al., 2004), bias-related racial trust disparity may potentially arise via affective responses to Black versus White faces. Our findings extend work finding that OFC activation to different groups reflects the increased liking of group members (Van Bavel et al., 2008),

Table 2

Covariate analyses showing individual differences in racial trust disparity predicted by neural response to Black > White faces.

Region	BA	k	t	MNI coordinates
<i>A. Increased activation with more racial trust disparity</i>				
R inferior parietal gyrus	40	17	3.44	54, 51, 54
<i>Increased activation with less racial trust disparity</i>				
L superior occipital gyrus	19	15	4.75	-18, -78, 21
R cerebellum (crus II)		20	4.04	51, -54, -45
R orbitofrontal cortex	10	18	3.38	18, 66, 0
<i>B. Increased connectivity with orbitofrontal cortex (18, 66, 0) with less racial trust disparity</i>				
<i>Occipital lobe</i>				
R fusiform gyrus	37/19	106	4.52	24, -51, -9
L primary visual cortex	17	77	4.44	-6, -75, -6
R fusiform gyrus	20	17	3.38	33, -27, -27
R inferior occipital gyrus	19	41	3.48	45, -78, -3
<i>Temporal lobe</i>				
R inferior temporal gyrus	20	15	4.41	63, -45, -24
L superior temporal gyrus	38	75	3.91	-51, 6, -12
R middle temporal gyrus	21	72	3.65	48, -42, -3
R superior temporal gyrus	48	55	3.60	66, -36, 24
L parahippocampal gyrus	28	20	3.58	-21, 9, -27
<i>Parietal lobe</i>				
L angular/superior parietal inferior parietal/postcentral gyrus/precuneus	39/7/40/2	1645	5.27	-24, -63, 33
L inferior parietal gyrus	2	41	4.06	51, -33, 54
R superior parietal gyrus	5	45	3.63	21, -57, 54
R postcentral gyrus	43	30	3.65	57, -9, 30
<i>Frontal lobe</i>				
L orbitofrontal gyrus	11	24	4.66	-18, 60, 9
L precentral gyrus	6	42	4.19	-48, 3, 57
L middle frontal gyrus	8	114	3.87	-27, 18, 51
R precentral gyrus	6	32	3.62	42, -9, 48
R middle frontal gyrus	46	15	3.59	48, 54, 0
L middle frontal gyrus	45	45	3.51	-45, 36, 36
R inferior frontal operculum	44	21	3.47	45, 21, 36
<i>Subcortical</i>				
R pallidum		60	4.23	15, -3, 6
R putamen		25	3.80	30, -18, 12
L caudate/putamen		32	3.70	-18, 6, 12
<i>Other</i>				
R cerebellum		71	3.79	42, -69, -24

by suggesting that more evaluation of Black relative to White faces may signal similar trustworthiness evaluations of these faces later. Supporting a role for OFC in trustworthiness evaluation, and reflecting a critical role for OFC in affective processing (Murray and Izquierdo, 2007), individuals with OFC lesions have difficulty using facial cues to make approachability judgments (Willis et al., 2010). Complementing this work, OFC activity uniquely decodes the perception of Black from White faces during affect-based friendship judgments, and the extent of its decoding correlates with evaluative (i.e., assessing race-valence associations) but not stereotyping (i.e., assessing race-stereotype associations) bias (Gilbert et al., 2012). Because participants were unaware they would evaluate trustworthiness after scanning, our findings uniquely show a robust relationship in how the brain perceives race and later explicit trustworthiness evaluations that potentially bypass cognitive control. This is a critical distinction because it suggests that individual differences in controlled processes do not affect all bias-related behaviors. Although participants were unaware of our focus on trustworthiness evaluations, however, control processes could have been engaged regardless given the fact that we presented only Black and White faces. This presentation may have raised participants' potential awareness of our interest in race, which resulted in increased engagement in race-related control regions.

Enhanced OFC-fusiform coupling when perceiving Black versus White faces also predicted less later racial trust disparity. A wide body of literature implicates the fusiform in perceptual face processing (Ishai et al., 2005), with responses decoding race in high, but not low,

Table 3

Covariate analyses showing individual differences in racial differentiation disparity predicted by neural response to Black > White faces.

Region	BA	k	t	MNI coordinates
<i>A. Increased activation with more differentiation disparity</i>				
R inferior occipital gyrus	19	23	4.07	33, -81, -9
L middle temporal gyrus	21	17	3.91	-45, -39, -3
R middle occipital gyrus	39	15	3.88	36, -66, 24
<i>Increased activation with less differentiation disparity</i>				
L ventrolateral prefrontal cortex	47	26	3.98	-30, 51, 3
R dorsolateral prefrontal cortex	46	40	3.84	24, 42, 21
R dorsolateral prefrontal cortex	46	18	3.33	27, 48, 15
<i>B. Increased connectivity with dorsolateral prefrontal seed (24, 42, 21) with less differentiation disparity</i>				
<i>Frontal lobe</i>				
L middle cingulate gyrus	6/24	54	5.04	-12, -18, 48
L rectus gyrus	11	27	5.01	0, 60, -18
L precentral gyrus	6	102	4.96	-27, -6, 45
R precentral gyrus	4	27	3.97	36, -27, 57
R precentral gyrus	6	47	3.41	51, -6, 48
<i>Parietal lobe</i>				
R postcentral gyrus	4	38	4.41	39, -27, 42
<i>Temporal lobe</i>				
R inferior temporal gyrus	20	36	3.37	51, -15, -18
L superior temporal gyrus	20	36	3.37	-51, -6, -9
R middle temporal gyrus	21	17	3.34	54, 3, -24
<i>Occipital lobe</i>				
L lingual gyrus	27	48	4.40	-15, -39, -3

bias perceivers (Brosch et al., 2012) and suggesting that less activity toward Black faces versus White begets their being processed as less “face-like” among White perceivers (Kubota et al., 2012). If OFC activity toward group members reflects affective processing in terms of liking (Van Bavel et al., 2008), coupling with regions implicated in enhanced perceptual processing of faces while merely perceiving race suggests that augmented perceptual-affective processing during race perception has critical implications for the later bias-related evaluations of faces in that they potentially reduce racial trust disparity. It is critical to note, however, that the racial trust disparity measured here reflects an *evaluative* race bias versus a bias based on stereotype application. Indeed, these differing implicit biases have distinct neural correlates (Amodio, 2014). Although we cannot rule out that demonstrated racial trust disparity resulted from stereotype application, the present results should be interpreted as a neural basis for individual differences in evaluative race biases.

Beyond revealing mechanisms for bias-related racial trust disparity, these data revealed that in the context of the same task, a unique neural signature predicts later race disparities unrelated to bias. Specifically, enhanced activity in dlPFC activity, a region widely identified in simple race perception tasks (Cunningham et al., 2004), linked to greater control of stereotypic responses to race (Amodio, 2010), and maintaining control more broadly (MacDonald et al., 2000) predicted less later racial differentiation disparity. People differentiate outgroup less than ingroup members (Linville et al., 1989), a finding replicated in our behavioral results. Differentiating between faces requires sufficient motivation and control (Hills and Lewis, 2006), suggesting that individual differences in regions implicated in both race perception and cognitive control should predict the relative differentiation of later trustworthiness evaluations of Black relative to White faces. Theoretical models of prejudice suggest that a first step in reducing prejudice is to foster more motivated individuated (i.e., differentiated) thinking about others (Fiske and Neuberg, 1990). Although bias and racial trust disparity were unrelated to racial differentiation disparity here, our neuroimaging results point to a possible intervention for future research to investigate that examines motivating participants to differentiate faces in order to potentially reduce overall racial trust disparity. Motivation to perceive people as individuals decreases the application of stereotypes to other-race faces (e.g., Blair, 2002). Motivation to perceive people as

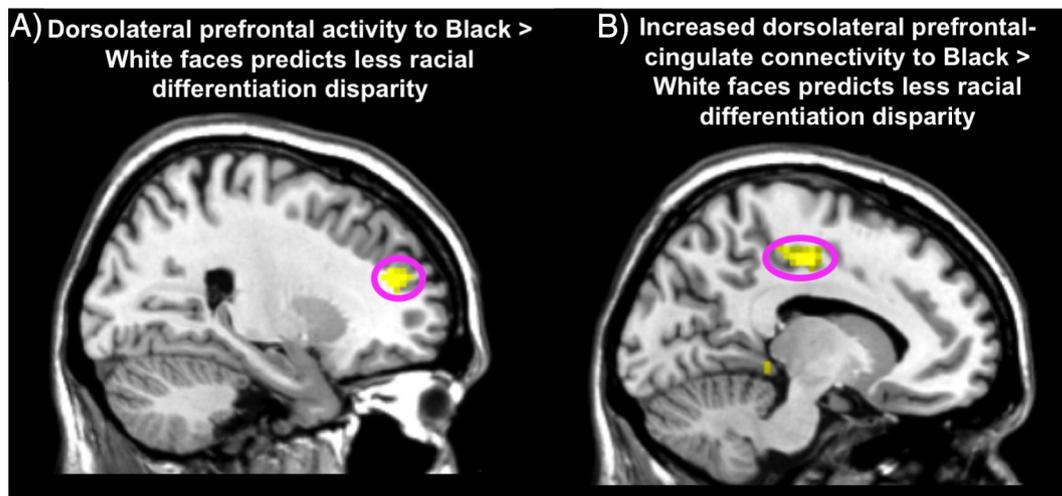


Fig. 2. The magnitude of racial differentiation disparity depended on individual differences in dorsolateral prefrontal response to Black versus White faces (A) and dorsolateral prefrontal-cingulate connectivity when perceiving Black versus White faces (B).

individuals may thus decrease *overall* racial disparity, although whether this would subvert the influence of implicit bias on disparity is unclear. Future research should directly test this intervention's effects on both behavior and associated neural response.

Functional coupling between dlPFC and the middle cingulate cortex when perceiving Black versus White faces predicted less racial differentiation disparity. Although we initially predicted coupling between dlPFC and other prefrontal regions typically implicated in cognitive control, coupling with cingulate may be tentatively interpreted within the context of prior work. The middle cingulate activates when individuals monitor changes in stimulus faces (Lee et al., 2012) and encode faces more broadly (Haxby et al., 1996; Ng et al., 2006), processes that contribute to individuation (Fiske and Neuberg, 1990). Future work should be designed to more closely examine a potential link between face differentiation and individuation-related processes reflected in neural activity. Of interest would be to determine whether a control-based mechanism predicting later racial differentiation disparity acts through individuation-related processing.

Curiously, the data did not evidence amygdala activation covarying with racial trust disparity as predicted. Prior work has found that the amygdala activates in the presence of very trustworthy and very untrustworthy faces (Said et al., 2009). At the same time, the amygdala also engages to Black versus White faces with very brief presentations (Cunningham et al., 2004) or responses that scale with implicit bias (Phelps et al., 2000). At first blush, it may seem curious that amygdala activity did not predict later racial trust disparity. Based on the literature, however, this may not have occurred for several reasons. First, that the Black and White faces in the present study did not differ in overall trustworthiness and that the amygdala tracks extreme instances of trustworthiness may have prohibited emergent amygdala response. Second, prior work finding race-related amygdala response in an event-related design has done so using very brief presentations, and not for stimulus durations similar to the 1-s duration used here (Cunningham et al., 2004). Third, while amygdala response to race may represent social group information via detection, it does not reflect race-based information that shapes trust decisions (Stanley et al., 2012). The absence of amygdala response does not mean that its activation does not contribute to racial trust disparity, however. While the present study offers initial insight into how brain activity corresponds with later racial trust disparity, future work should examine a more exact time course of this relationship to further elucidate neural contributions to behavior.

The present task found evidence for unique neural signatures within the race perception network predicting bias-related racial trust

disparity and racial differentiation disparity obtained from the same task. However, it is important to note that these findings should be kept in the context of race disparities related to aspects of trustworthiness ratings. For instance, just because dlPFC activity contributed to racial differentiation disparity and was unassociated to bias-related race disparities here, it does not suggest that dlPFC activity never predicts bias-related behaviors. Illustrating the importance of task context in interpreting race disparities, prior work on interracial contact shows that individuals with higher versus lower levels of implicit bias have more interference on the Stroop task, which draws upon controlled processing, after participating in interracial interactions (Richeson and Shelton, 2003). Future studies may show that different components of the race perception network are sensitive to bias or not depending on task context. However, the present work critically shows that dissociable aspects of the race perception network underlie later obtained in bias-related and differentiation-based race disparities associated with trustworthiness. These dissociations have critical implications for the development of strategies targeted to specifically reduce racial trust disparity and racial differentiation disparity.

It is also important to note that although the a priori regions of interest discussed here (i.e., OFC, fusiform, and dlPFC) are well-characterized in terms of race perception (Amodio, 2014), they are also implicated in processes beyond those discussed here. To justify the interpretation of these findings, we verified that the locations of emergent regions were involved in face processing, affective processing, and cognitive control, respectively, via a large database of neuroimaging studies (Yarkoni et al., 2011). We specifically verified that the regions activated in the present task had also been activated by tasks manipulating face processing, affective processing, and cognitive control. This suggests that interpreting our findings in the context of perceptual, affective, and cognitive processing is appropriate. However, future work on race and trustworthiness evaluation can verify the nature of the present findings by manipulating these processes and by studying the effects of these manipulations on the relationship between brain activity and later aspects of trustworthiness evaluation. It would also be worthwhile to examine whether these contributions to later behavior are causal by testing the relationship of different activation patterns to online trustworthiness evaluations.

Trustworthiness evaluations and their differentiation among racial minorities have critical implications for intergroup interactions, particularly for cross-racial empathy. Indeed, prior ERP work has found disparity in components of empathy by race (Sessa et al., 2014). Critically, increased other-race contact reduces the association of racial bias with race disparities in empathy (Avenanti et al., 2010; Cao et al., 2015),

and also impacts neural responses to race (Cloutier et al., 2014). Because people have stronger empathic responses to those who look trustworthy (Sessa and Meconi, 2015), increased other-race contact may potentially alter neural response to race in a way that decreases racial trust disparity and also reduces race disparity in empathy. By revealing distinct neural mechanisms that support later aspects of disparities in trustworthiness evaluation, the present data enhance our knowledge of the neural correlates of race perception and related behavior. The understanding gained from these findings can be leveraged by future work to reduce the pernicious societal consequences of race disparities in social cognition.

Acknowledgments

We thank Bryan Bennett and Gregory Sprout for research assistance.

References

- Amodio, D., 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. *J. Cogn. Neurosci.* 22 (11), 2609–2617.
- Amodio, D., 2014. The neuroscience of prejudice and stereotyping. *Nat. Rev. Neurosci.*
- Amodio, D., Devine, P., 2006. Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *J. Pers. Soc. Psychol.* 91 (4), 652–661.
- Avenanti, A., Sirigu, A., Aglioti, S., 2010. Racial bias reduces empathic sensorimotor resonance with other-race pain. *Curr. Biol.* 20 (11), 1018–1022.
- Beer, J., Stallen, M., Lombardo, M., Gonsalkorale, K., Cunningham, W., Sherman, J., 2008. The quadruple process model approach to examining the neural underpinnings of prejudice. *NeuroImage* 43, 775–783.
- Blair, I., 2002. The malleability of automatic stereotypes and prejudice. *Personal. Soc. Psychol. Rev.* 6 (3), 242–261.
- Brigham, J., 1993. College students' racial attitudes. *J. Appl. Psychol.* 23, 1933–1967.
- Brosch, T., Bar-David, E., Phelps, E., 2012. Implicit race bias decreases the similarity of neural representations of Black and White faces. *Psychol. Sci.* 24 (2), 160–166.
- Cao, Y., Contreras-Huerta, L., McFadyen, J., Cunningham, R., 2015. Racial bias in neural response to others' pain is reduced with other-race contact. *Cortex* 70, 68–78.
- Chekroud, A., Everett, J., Bridge, H., Hewstone, M., 2014. A review of neuroimaging studies of race-related prejudice: does amygdala response reflect threat? *Front. Hum. Neurosci.*
- Cloutier, J., Li, T., Correll, J., 2014. The impact of childhood experience on amygdala response to perceptually familiar black and white faces. *J. Cogn. Neurosci.* 26 (9), 1992–2004.
- Cunningham, W., Johnson, M., Raye, C., Gatenby, J., Gore, J., Banaji, M., 2004. Separable neural components in the processing of black and white faces. *Psychol. Sci.* 15 (12), 806–813.
- DeSteno, D., Dasgupta, N., Bartlett, M., Caidric, A., 2004. Prejudice from thin air: the effect of emotion on intergroup attitudes. *Psychol. Sci.* 15 (5), 319–324.
- Devine, P., 1989. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18.
- Dotsch, R., Wigboldus, D., Langner, O., van Knippenberg, A., 2008. Ethnic out-group faces are biased in the prejudiced mind. *Psychol. Sci.* 19 (10), 978–980.
- Dovidio, J., Brigham, J., Johnson, B., Gaertner, S., 1996. Stereotyping, prejudice, and discrimination: another look. In: Macrae, C., Stangor, C., Hewstone, M. (Eds.), *Stereotypes and stereotyping*. Guilford Press, New York, NY, pp. 276–292.
- Feng, L., Liu, J., Wang, Z., Li, J., Li, L., Ge, L., ... Lee, K., 2011. The other face of the other-race effect: an fMRI investigation of the other-race face categorization advantage. *Neuropsychologia* 49 (13), 3739–3749.
- Ferguson, D., Rhodes, G., Lee, K., Sriram, N., 2001. 'The all look alike to me': prejudice and cross-race face recognition. *Br. J. Psychol.* 92 (4), 567–577.
- Fiske, S., Neuberg, S., 1990. A continuum model of impression formation from category-based to individuating processes: influences of information and motivation on attention and interpretation. *Adv. Exp. Soc. Psychol.* 23, 1–74.
- Gawronski, B., Geschke, D., Banse, R., 2003. Implicit bias in impression formation: associations influence the construal of individuating information. *Eur. J. Soc. Psychol.* 33, 573–589.
- Gilbert, S., Swencionis, J., Amodio, D., 2012. Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia* 50, 3600–3611.
- Golby, A., Gabriell, Chiao, J., Eberhardt, J., 2001. Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845–850.
- Goldstein, A., Chance, J., 1985. Effects of training on Japanese face recognition: reduction of the other-race effect. *Bull. Psychon. Soc.* 23, 211–214.
- Greenhaus, J., Parasuraman, S., Wormley, W., 1990. Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Acad. Manag. J.* 33 (1), 64–86.
- Greenwald, A., Banaji, M., 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *J. Pers. Soc. Psychol.* 102, 4–27.
- Greenwald, A., Nosek, B., Banaji, M., 2003. Understanding and using the Implicit Association Test: an improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216.
- Hart, A., Whalen, P., Shin, L., McInerney, S., Fischer, H., Rauch, S., 2000. Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroReport* 11 (11), 2351–2354.
- Haxby, J., Ungerleider, L., Horwitz, B., Maisog, J., Rapoport, S., Grady, C., 1996. Face encoding and recognition in the human brain. *Proc. Natl. Acad. Sci.* 93, 922–927.
- Hills, P., & Lewis, M. (2006). Reducing the own-race bias in face recognition by shifting attention. *Q. J. Exp. Psychol.* 59(6), 2006.
- Ishai, A., Schmidt, C., Boesiger, P., 2005. Face perception is mediated by a distributed cortical network. *Brain Res. Bull.* 67, 87–93.
- Kanwisher, N., McDermott, J., Chun, M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17 (11), 4302–4311.
- Kringelbach, M., Rolls, E., 2004. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Process Neurobiol.* 72 (5), 341–372.
- Kubota, J., Banaji, M., Phelps, E., 2012. The neuroscience of race. *Nat. Neurosci.* 15 (7), 940–948.
- Kubota, J., Li, J., Bar-David, E., Banaji, M., Phelps, E., 2013. The price of racial bias: intergroup negotiations in the Ultimatum Game. *Psychol. Sci.* 24 (12), 2498–2504.
- Lee, Y., Anaki, D., Grady, C., Moscovitch, M., 2012. Neural correlates of temporal integration in face recognition. *NeuroImage* 61 (4), 1287–1299.
- Li, J., Dunning, D., Malpass, R., 1998. Cross-racial identification among European-Americans: basketball fandom and the contact hypothesis. Paper presented at the American Psychology-Law Society, Redondo Beach, CA.
- Linville, P., Fischer, G., 1998. Group variability and covariation: effects of intergroup judgment and behavior. In: Sedikides, C., Schopler, J., Insko, C. (Eds.), *Intergroup cognition and intergroup behavior*. Erlbaum, Hillsdale, NJ, pp. 123–150.
- Linville, P., Fischer, G., Salovey, P., 1989. Perceived distributions of the characteristics of in-group and out-group members: empirical evidence and a computer simulation. *J. Pers. Soc. Psychol.* 57 (2), 165–188.
- MacDonald, A., Cohen, J., Stenger, V., Carter, C., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288 (5472), 1835–1838.
- McConnell, A., Leibold, J., 2001. Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *J. Exp. Soc. Psychol.* 37 (5), 435–442.
- McLaren, D., Ries, M., Xu, G., Johnson, S., 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage* 61 (4), 1277–1286.
- Meconi, F., Luria, R., Sessa, P., 2014. Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Soc. Cogn. Affect. Neurosci.* 9 (12), 1872–1879.
- Meissner, C., Brigham, J., 2001. Thirty years of investigating the own-race bias in memory for faces. *Psychol. Public Policy Law* 7, 3–35.
- Murray, E., Izquierdo, A., 2007. Orbitofrontal cortex and amygdala contributions to affect and action in primates. *Ann. N. Y. Acad. Sci.* 1121, 273–296.
- Ng, M., Ciaramitaro, V., Anstis, S., Boynton, G., Fine, I., 2006. Selectivity for the configural cues that identify the gender, ethnicity, and identify of faces in human cortex. *Proc. Natl. Acad. Sci.* 103, 19552–19557.
- Phelps, E., O'Connor, D., Cunningham, W., Funayama, E., Gatenby, J., Gore, J., Banaji, M., 2000. Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12 (5), 729–738.
- Plant, E., Devine, P., 1998. Internal and external motivation to respond without prejudice. *J. Pers. Soc. Psychol.* 75 (3), 811–832.
- Richeson, J., Shelton, J., 2003. When prejudice does not pay: effects of interracial contact on executive function. *Psychol. Sci.* 14 (3), 287–290.
- Richeson, J., Baird, A., Gordon, H., Heatherton, T., Wyland, C., Trawalter, S., Shelton, J., 2003. An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* 6 (12), 1323–1328.
- Ronquillo, J., Denson, T., Lickel, B., Lu, Z., Nandy, A., Maddox, K., 2007. The effects of skin tone on race-related amygdala activity: an fMRI investigation. *Soc. Cogn. Affect. Neurosci.* 2 (1), 39–44.
- Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. *Behav. Neurol.* 12 (4), 191–200.
- Rule, N.O., Krendl, A.C., Ivcevic, Z., Ambady, N., 2013. Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *J. Pers. Soc. Psychol.* 104 (3), 409.
- Said, C., Baron, S., Todorov, A., 2009. Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *J. Cogn. Neurosci.* 21 (3), 519–528.
- Sessa, P., Meconi, F., 2015. Perceived trustworthiness shapes neural empathic responses toward others' pain. *Neuropsychologia* 79A, 97–105.
- Sessa, P., Meconi, F., Castelli, L., Dall'Acqua, R., 2014. Taking one's time in feeling other-race pain: an event-related potential investigation on the time-course of cross-racial empathy. *Soc. Cogn. Affect. Neurosci.* 9 (4), 454–463.
- Stone, A., Brigham, J., Meissner, C., 2000. Social and cognitive factors affecting the own-race bias in Whites. *Basic Appl. Soc. Psychol.* 22, 71–84.
- Slotnick, S., Moo, L., Segal, J., Hart, J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cogn. Brain Res.* 17, 75–82.
- Stanley, D., Sokol-Hessner, P., Banaji, M., Phelps, E., 2011. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proc. Natl. Acad. Sci.* 108 (19), 7710–7715.
- Stanley, D., Sokol-Hessner, P., Fareri, D., Perino, M., Delgado, M., Banaji, M., Phelps, E., 2012. Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philos. Trans. R. Soc. B* 367 (744-753).

- Todorov, A., 2008. Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Ann. N. Y. Acad. Sci.* 1124, 208–224.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15 (1), 273–289.
- Van Bavel, J., Packer, D., Cunningham, W., 2008. The neural substrates of in-group bias: a functional magnetic imaging study. *Psychol. Sci.* 11, 1131–1139.
- Vuilleumier, P., Armony, J., Driver, J., Dolan, R., 2001. Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron* 30 (3), 829–841.
- Wheeler, M., Fiske, S., 2005. Controlling racial prejudice: social-cognitive goals affect amygdala and stereotype activation. *Psychol. Sci.* 16 (1), 56–63.
- Willis, M., Palermo, R., Burke, D., McGrillen, K., Miller, L., 2010. Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. *Neuropsychologia* 48 (7), 2182–2187.
- Winston, J., Strange, B., O'Doherty, J., Dolan, R., 2002. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5 (3), 277–283.
- Yarkoni, T., Poldrack, R., Nichols, T., Van Essen, D., Wager, T., 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8 (8), 665–670.