Political Psychology

Political Psychology, Vol. xx, No. xx, 2012
doi: 10.1111/j.1467-9221.2012.00879.x

Huxtables on the Brain: An fMRI Study of Race and Norm Violation

Darren Schreiber
University of California, San Diego

Marco Iacoboni
University of California, Los Angeles

While a substantial body of work has been devoted to understanding the role of negative stereotypes in racial attitudes, far less is known about how we deal with contradictions of those stereotypes. This article uses functional brain imaging with contextually rich visual stimuli to explore the neural mechanisms that are involved in cognition about social norms and race. We present evidence that racial stereotypes are more about the stereotypes than about race per se. Amygdala activity (correlated with negative racial attitudes in other studies) appeared driven by norm violation, rather than race. Similarly, a pattern of deactivation in the medial prefrontal cortex (previously associated with the dehumanizing of social outcasts) was connected to norm violation, not race.

KEY WORDS: race, social norms, neuropolitics, brain imaging, fMRI

Introduction

On the sitcom *The Cosby Show*, Bill Cosby portrayed obstetrician Cliff Huxtable, married to attorney Clair Huxtable, and depicted a successful African American family living the American dream. While a substantial body of work has been devoted to understanding the role of negative stereotypes in racial attitudes, far less is known about how we deal with contradictions of those stereotypes. Do members of other racial groups remain in the outgroup, even when they are exemplars of mainstream social norms? Or, are negative racial attitudes the result of implicit stereotypes that are defeated when explicit positive contexts are encountered? Finally, what neural mechanisms are engaged when people process race and social norm violation?

The violation of social norms is at the heart of the very need for government.[1] However, we have few insights into the mechanisms through which we detect and process violations of those social norms. Here, we use functional brain imaging to investigate the brain regions involved when participants perceive images of African Americans and Europeans Americans who are either consistent with social norms or violating them. We find evidence that it is racial stereotypes that are

---

[1] "If men were angels, no government would be necessary. If angels were to govern men, neither external nor internal controls on government would be necessary." Federalist 51.

0162-895X © 2012 International Society of Political Psychology
Published by Wiley Periodicals, Inc., 350 Main Street, Malden, MA 02148, USA, 9600 Garsington Road, Oxford, OX4 2DQ,
and PO Box 378 Carlton South, 3053 Victoria, Australia

activating the amygdala, a brain region often thought of as being involved in threat perception, rather than race itself. Similarly, we find that it is violators of social norms that provoke the dehumanizing reactions evidenced in the pattern of deactivation in the medial prefrontal region of the brain, while both norm-consistent African American and European Americans trigger activation, suggesting contemplation of the mental processes of others. Finally, we present some results that may pave the way for a more nuanced understanding of the social cognition of race, further investigation into the mechanisms behind social norms, and suggest new lines of inquiry for understanding populations with pathological tendencies to violate social norms, such as patients with borderline personality disorder.

Theory

Active debates rage about whether racial tensions are driven by principled conservatism (Sniderman & Carmines, 1997), symbolic racism (Sears, Van Laar, Carrillo, & Kosterman, 1997), evolutionary mechanisms (Kurzban, Tooby, & Cosmides, 2001), real conflict in group interests (Bobo, 1983), or some mix of these (Sidanius & Pratto, 1999). A central theme in many of these debates is the role of social norms and whether these exist as a reflection of racial divisions or as a cause of them. For instance, is opposition to welfare policies driven by seeing the poor as a social outgroup in themselves or by stereotyping the poor as being black (Gilens, 1999)?

Disentangling the components of racial attitudes is made more difficult because of the wide variety of mental mechanisms at play. Research has shown that implicit and explicit racial attitudes have distinct characteristics and implications (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Greenwald, Poehlman, Uhlmann, & Banaji, 2009) and rely on different neural mechanisms (Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005; Phelps et al., 2000). Racial attitudes can also vary with motivations (Wheeler & Fiske, 2005) to see the racial others as part of an ingroup or an outgroup (Kurzban et al., 2001; Van Bavel & Cunningham, 2009; Van Bavel, Packer, & Cunningham, 2008) and depending on whether one is able to reflect upon the racial other as an individual or as merely part of stereotyped outgroup (Freeman, Schiller, Rule, & Ambady, 2010). People also demonstrate a tendency to mispredict how they will respond when encountering racist attitudes (Kawakami, Dunn, Karmali, & Dovidio, 2009). That poor prediction may be due to self-deception or the strong social desirability effects of a culture that reinforces egalitarian public declarations.

Functional brain imaging provides an important opportunity to distinguish the mental processes that underpin racial attitudes and behavior (Eberhardt, 2005; Ito & Bartholow, 2009). Researchers use functional magnetic resonance imaging (fMRI) to investigate changes in blood flow to different regions of the brain with the goal of identifying specific neural systems engaged by mental processes. The blood oxygen level dependent (BOLD) fMRI signal fluctuates with alterations in the ratio of oxygenated to deoxygenated hemoglobin. It is believed that the additional work that neurons are doing to support a particular mental task drives an influx of oxygenated blood to provide energy for that additional work. Thus, the analysis of changes in the BOLD fMRI signal allows us to make inferences about shifts in cerebral blood flow, which allows us to infer the activity of neurons in particular regions of the brain.

One of the paradoxes of fMRI research is that it often attempts to localize mental functions to particular brain regions, while we also know that the brain is phenomenally able to reorganize in response to experience or trauma. For instance, learning a new skill can alter both the thickness of the neocortex in specific regions (Draganski et al., 2004) and change structural connectivity among regions (Scholz, Klein, Behrens, & Johansen-Berg, 2009). Nonetheless, meta-analysis reveals remarkable consistency in connections between mental function and regional brain activation (Costafreda, Brammer, David, & Fu, 2008; Van Overwalle, 2009).

Because there is a demonstrated tendency to misunderstand the meaning of biological correlates of social behavior, we want to make clear at the outset that we are not making claims about the innateness of particular predispositions or taking a position about whether biology *determines* social behavior. Rather, this work is premised on the idea that mental processes are reflected in the brain. Even the original Cartesian dualist believed that there was a connection between mental processes and the brain (Descartes, 1641/1979). Thus, our study seeks to investigate the specific connections between particular mental tasks (e.g., processing race and social norms) and the brain regions that support those mental tasks.

Race is a particularly promising area for investigating with functional brain imaging because of the limitations that traditional tools of political science like surveys face. Phenomena like the race of interviewer effect (Hatchett & Schuman, 1975), stereotype threat (Davis & Silver, 2003), social desirability (Carver, Glass, & Katz, 1978), and implicit attitudes (Greenwald et al., 2009) each impair the interpretability of survey results on racial attitudes in complex ways.[2] Survey experiments (Sniderman & Carmines, 1997), a variety of laboratory experimental techniques (Hewstone, Rubin, & Willis, 2002), and field experiments (Paluck & Green, 2009) have supplemented the insights gained in surveys. However, identifying particular brain regions activated can further constrain potential theories.

While a few dozen fMRI experiments on race have been done and while there are many consistent findings, one of the most significant limitations (and one illustrated by the findings in this article) is on the interpretability of the regional brain activity that is detected. Though the amygdala is canonically described as involved in the processing of threats, it has also been shown to be involved in a wide range of social emotional processing, including positive emotions (Costafreda et al., 2008; Sergerie, Chochol, & Armony, 2008). Nonetheless, this association between negative affect and racial outgroups has been the focus of much of the functional imaging work in this area.

Early work identified the amygdala (a pair of almond sized structures located adjacent to the brain stem; see Figures 1 and 2) as active while perceiving images of members of racial outgroups (Hart et al., 2000) and tied the intensity of amygdala activity to other measures of negative implicit associations connected to members of racial outgroups (Phelps et al., 2000). However, damage to the

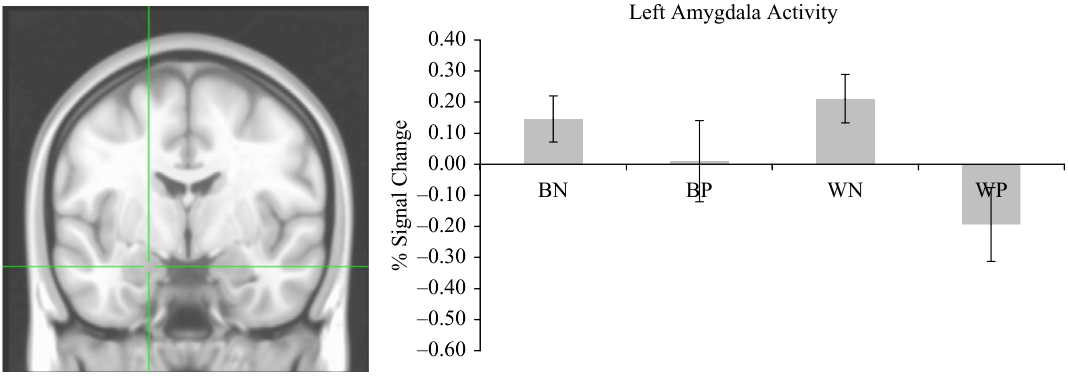

Figure 1. Activity in the left amygdala corresponding with viewing images of norm-violating African Americans (BN), norm-consistent African Americans (BP), norm-violating European Americans (WN), and norm-consistent European Americans (WP).

---

[2] For a recent and excellent review of the variety of methodological and theoretic approaches to racial attitudes, see Huddy and Feldman (2009).

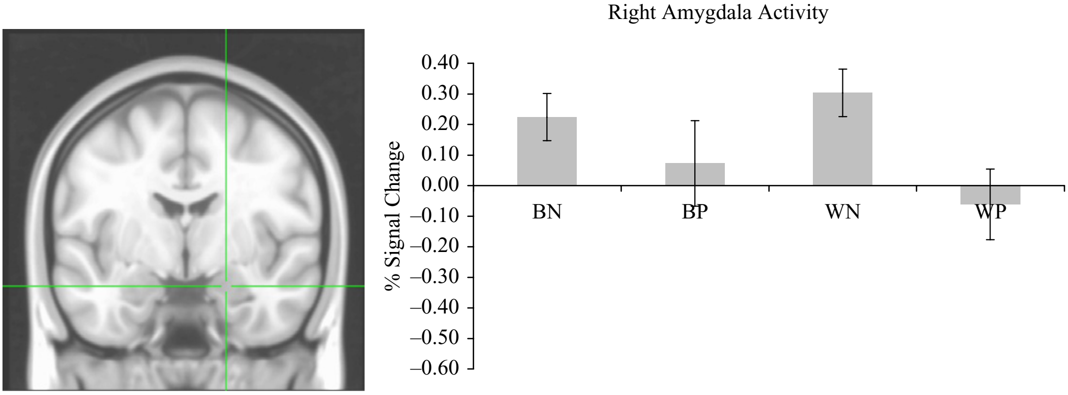

Figure 2. Activity in the right amygdala corresponding with viewing images of norm-violating African Americans (BN), norm-consistent African Americans (BP), norm-violating European Americans (WN), and norm-consistent European Americans (WP).

amygdala did not appear to impair typical performances on the implicit association task or discrepancies between explicit and implicit racial attitudes (Phelps, Cannistraci, & Cunningham, 2003). The amygdala did appear more responsive to subliminally presented stimuli than to stimuli that were presented for a duration long enough to be consciously perceived (Cunningham et al., 2004). The differences between automatic/implicit and controlled/explicit racial attitudes appeared to be tied to differences in amygdala activation as modulated by the frontal lobes (Lieberman et al., 2005).

However, the amygdala is not the only brain region implicated in the processing of racial phenomena. The fusiform face area and medial temporal lobe have both been shown to play a role in our ability to recall the faces of members of racial outgroups (Golby, Gabrieli, Chiao, & Eberhardt, 2001). The anterior cingulate cortex's empathic response to the physically (Xu, Zuo, Wang, & Han, 2009) and socially (Krill & Platek, 2009) inflicted pain of others is diminished substantially when the other is of another race. Ingroup/outgroup effects have also been implicated in the default state network's processing of a social cooperation task (Rilling, Dagenais, Goldsmith, Glenn, & Pagnoni, 2008). This network includes the medial prefrontal cortex, which has been shown to be involved in considering the mental states of others. However, when the other is a member of an undesirable social outgroup, the mentalizing activity in this region diminishes (Harris & Fiske, 2006).

The generalizability of many of these results beyond the American black/white dimension that dominated much of the early work in the field supports a recognition of the role of a wider and more nuanced range of ingroup/outgroup dynamics. Theoretical frames like social dominance orientation (Sidanius & Pratto, 1999) or evolutionary psychology (Kurzban et al., 2001) suggest that contemporary racial attitudes are perhaps a particular instance of a more basic tendency to form and navigate social alliances. These frames are supported by work showing that the same brain regions involved in processing race are also involved when other ingroup/outgroup distinctions are made salient (Van Bavel et al., 2008).

The Experiment

Thus, we designed this experiment to contrast two different types of ingroup/outgroup distinctions and see if we could differentiate the brain regions that were involved in the distinctions. We recruited a group of 19 college students to view a set of images while they were being scanned using fMRI. The subjects were recruited from the college Republican club, college Democrat club, or from a sample of volunteers who had low scores on a measure of political knowledge. The images they viewed were of either European Americans or African Americans who were either norm-violating (e.g., criminals, gang members, homeless) or norm consistent (e.g., families, students, doctors; see the Appendix for a detailed description of the participants, the stimuli, and the data analysis).

The participants viewed the images in block design where five images of one type (e.g., European American norm violators) were shown one after another for 4 seconds each. After each block, the subjects would have 20 seconds of watching a blank screen. The functional brain images acquired while the subjects participated in the experiment were then registered onto a structural MRI of the individual's brain and those individual brains were registered into a standard brain atlas. In the statistical analysis, we used a general linear model to estimate the fit at each voxel (the 3D equivalent of a pixel).

There are two main approaches that are used in analyzing fMRI data: region of interest (ROI) analysis and whole brain analysis. Because the prior literature indicated that the amygdala was frequently activated when people viewed images of African Americans, we extracted the data for both the left and right amygdala and compared it with a resting baseline. Based on prior results, we expected that the amygdala would be active for both the norm-violating and norm-consistent images of African Americans, but also might be active with European American norm violators.

Other work (Freeman et al., 2010; Harris & Fiske, 2006) has suggested that activity in an ROI in the medial prefrontal cortex (mPFC) might be diminished while viewing social outgroups. The mPFC is located directly behind the middle of the forehead and is involved, among other thing, in contemplating the mental states of others (see Amodio & Frith, 2006 for a review). The region is typically at a fairly high level of metabolic activity even during rest, but it is part of a network of brain regions that deactivates during many tasks of technical cognition (Gusnard & Raichle, 2001; Northoff, Qin, & Nakao, 2010; Raichle et al., 2001). A wide variety of tasks of social cognition cause activity here to increase above the resting baseline (Schreiber, 2011). Dysfunction in the area is implicated in autism, which is typically characterized by difficulty with reading the intentional states of others (Iacoboni, 2006; Minshew & Keller, 2010). While activations here are often found in competitive or cooperative contexts when we engage with other humans, the region does not activate in parallel contexts when the other player is a computer (McCabe, Houser, Ryan, Smith, & Trouard, 2001). Thus, patterns of deactivations in the mPFC have been interpreted as potentially a form of "dehumanization" wherein the individual thinks of an other as an "it" rather than another human of equal social cognitive status. The ROI analysis allows us to test specific hypotheses about the activity of regions during each of the experimental conditions. The whole brain analysis, in contrast, tests a model against all of the voxels in the entire brain at once, which requires appropriate statistical corrections.

Results

Region of Interest Analysis

The ROI analysis for the amygdala proved to be largely consistent with the prior literature. Images of African Americans appeared to activate both the left and the right amygdala. Similarly, the norm-violating images activated the amygdala bilaterally, while norm-consistent images deactivated it.

However, decomposing the results further supports the implication of the implicit association task literature that it is negative stereotypes of African Americans that are driving the amygdala results when there is no social context provided for the faces. Norm-consistent images of African Americans neither significantly activate nor deactivate the amygdala, whereas norm-violating images of European Americans do activate it (Figures 1 and 2). It may well be that presenting the disembodied face of an African American, as in a typical social cognitive neuroscience experiment, triggers a network of associations with social violations that is reinforced by negative stereotypic

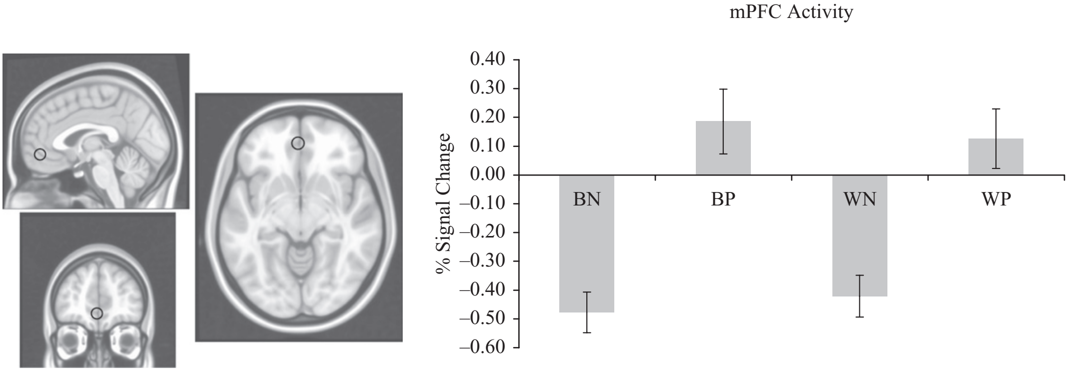

Figure 3. Activity from a region of interest in the medial prefrontal cortex (–2, 48, –7) corresponding with viewing images of norm-violating African Americans (BN), norm-consistent African Americans (BP), norm-violating European Americans (WN), and norm-consistent European Americans (WP).

images of African Americans presented in the media. When presented with an image of African Americans that explicitly reflects the negative stereotypes, the amygdala is activated. However, presenting images that run contrary to the negative stereotype dampens the typical associations and thus may preclude the amygdala from activating.

In contrast, typical imaging studies on race find deactivation in the amygdala when participants view disembodied faces of European Americans. The results here suggest that this may also be the result of implicit associations. When the European Americans are shown in a norm-consistent context, the left amygdala deactivates in a statistically significant manner (Figure 1). But, when they are shown as norm violators, one obtains the same level of activation as for African American norm violators (Figures 1 and 2).

These results suggest an interpretation akin to Sniderman and Carmines' (1997) principled conservativism. They used a survey experiment in which questions pertaining to social policies such as immigration, welfare, and affirmative action were randomly altered to include references to African Americans or white Europeans, the deserving or the undeserving, and based on various principles. Sniderman and Carmines contend that it is the violation of conservative principles (i.e., a specific set of social norms) that triggers opposition to these policies, rather than race per se. The region of interest analysis here lends credence to this perspective. Participants' amygdala activity to African American faces in other studies appears driven by the implicit associations with norm violation that come with racial stereotypes in America. When those associations are explicitly violated, then the purported threat response measured in the amygdala evaporates.

The results with the ROI in the mPFC show a similar pattern. The norm-violating images deactivate this region, while the norm-consistent images activate it (Figure 3). Decomposing these results shows that the norm-violating African Americans provoke the largest deactivation, while the norm-consistent African Americans have the largest activation. Thus, norm violation (for both African American and European American images) diminishes activity in this region associated with mentalizing, and norm-consistent images increase the apparent mentalizing activity.

Thus again, it is norm violation that appears to be driving the differences in mPFC, rather than merely race. Harris and Fiske (2006) argue that it is those appraised to be low in competence and low in social warmth that are the "lowest of the low" and thus are dehumanized (as reflected in diminished mPFC). The failure of individuated African Americans to activate the mPFC as individuated European Americans do (Freeman et al., 2010) may have been the result of disembodied faces that were not sufficiently contextualized by the handful of narrative descriptors. With a rich

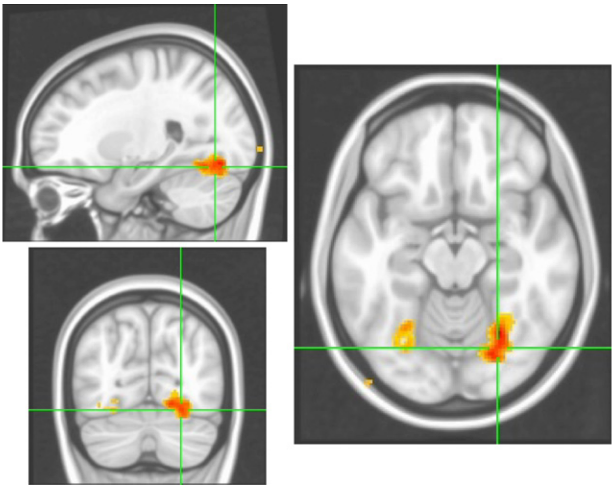

Figure 4. Bilateral activity in the fusiform gyrus (crosshairs located at (24, −72, −21)) for a contrast of norm-violating versus norm-consistent images.

visual context, both norm-violating African Americans and European Americans deactivate the mPFC and norm-consistent images of both groups activate it.

Both sets of ROI results are consistent with previous imaging studies in the area of race, but put the emphasis on ingroup/outgroup effects, rather than on race per se. While negative stereotypes may implicitly put African Americans in a default outgroup status, the effect of those stereotypes on the brain appear to be overridden with relative ease. We may perceive norm-violating African Americans as threats and dehumanize them, but we do the same with norm-violating European Americans.

Whole Brain Analysis

The results of the whole brain analysis were much more of a surprise (Figure 4, Table 1). With the more severe statistical thresholding applied to the model run over the entire volume of the brain, we found no clusters of activation large enough to meet our standard for the contrasts based on race. The contrast for norm violation versus norm-consistent images showed bilateral activations in the fusiform gyrus. Decomposing the results demonstrated activation in the fusiform for both African American and European American norm-violating images. Furthermore, deactivations were found in both African American and European American norm-consistent images.

Discussion

These large bilateral activations in the fusiform gyrus were initially puzzling. The fusiform is located at the bottom of the brain, at the intersection of the temporal and occipital cortices, and is most often discussed in terms of its role in higher level processing of visual information, specifically the recognition and categorization of objects such as cars, birds, and faces (Gauthier, Skudlarski, Gore, & Anderson, 2000). This categorization activity can be driven by both top-down and bottom-up cognitive processes (Devlin, Rushworth, & Matthews, 2005). The amygdala is typically considered to be part of a bottom-up processing stream (Lieberman, Schreiber, & Ochsner, 2003) and its responsiveness to social anxiety and fear is moderated by activity in the fusiform (Pujol et al., 2009). The amygdalo-fusiform pathway of neural connections runs parallel

Table 1. Clusters of Activation for a Contrast of Norm-Violating Versus Norm-Consistent Images

Voxels	z-score	x	y	z	Anatomy
729	4.19	16	−74	−10	Right Fusiform
359	3.61	−30	−100	6	Left Middle Occipital
73	3.54	36	−80	12	Right Middle Occipital
63	3.01	−26	−58	−14	Left Fusiform
59	3.33	18	−56	14	Right Posterior Cingulate

Note. Thresholded for $z > 2.3$ (corrected) with clusters having a spatial extent of 50 voxels or more (Worsley, 2001).

to the hippocampo-fusiform pathway (Smith et al., 2009), with the hippocampus typically being considered as part of a top-down processing stream (Lieberman et al., 2003). These parallel networks may provide insight into a potential social function of the fusiform, especially due to the importance of the amygdala for a variety of social cognition tasks.

It is the role of this region in the perception of faces that has garnered the most attention and controversy. Some have proposed the existence of a fusiform face area (FFA) that is specifically focused on processing faces (Kanwisher, McDermott, & Chun, 1997). This region is usually identified by a functional localizer task that identifies a region of the fusiform that responds specifically to faces and is less responsive to other types of objects (Berman et al., 2010). The FFA has been so widely discussed and debated that many studies involving social cognition attribute activations in the fusiform to face processing regardless of whether they have appropriately localized the FFA.

While there are faces present in all of the stimuli in this experiment, the pattern of activity here was much larger in volume than in the typical face processing experiments, and there were no clear precedents that would explain the pattern of activation and deactivation. The fusiform does appear to respond relative to conscious attention (Kastner, De Weerd, Desimone, & Ungerleider, 1998; Reddy, Moradi, & Koch, 2007; Wojciulik, Kanwisher, & Driver, 1998), and it is possible that the norm-violating stimuli would attract more attention as the participants scan a potentially more threatening scene. Furthermore, attention to fearful faces can activate the fusiform (Vuilleumier, Armony, Driver, & Dolan, 2001), while inattention to those faces will diminish the activity (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002). The fusiform is also active when subjects attend to the shapes (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990) or surface properties (Cant & Goodale, 2007) of objects.

Thus, one possibility is that scenes of potential social norm violation draw more attention to the objects present in that scene. In this scenario, lower-level visual processes send information to higher-level processes, which in turn focus the effort of those visual regions on potential evidence of social norm violation. This would be akin to what happens when subjects were asked to attend to particular features of objects. Such asymmetric attention to norm-violating versus norm-supporting images would be consistent with Prospect Theory (Kahneman & Tversky, 1979) and its prediction of asymmetric sensitivity to potential losses. If attention was driving the differences then we might also expect to see other brain regions (Bush & Shin, 2006) that are classically involved in attention to be also activated, which we do not. Another possibility is that the fusiform itself plays a more direct role in the detection of norm violation and thus drives the higher-level processes.

Intriguingly, people with Borderline Personality Disorder (BPD) have difficulties with processing affect in faces and with processing social norms, presenting an interesting possibility. BPD is a psychiatric disorder with significant implications for social dynamics and is found in about 1.8% of the population. The pop culture image of BPD is Glenn Close's character in the film *Fatal Attraction*, a woman with intense mood swings who tends to see people in dramatic black or white terms. The social costs of the disease are substantial, with extremely high rates of suicide, and account for a significant portion of psychiatric hospitalization (Lieb, Zanarini, Schmahl, Linehan, & Bohus, 2004).

People with BPD often struggle with recognizing when they have broken a social norm and are unable to repair a relationship once it has begun to deteriorate. In a repeated trust game, for instance, healthy controls are able to maintain high levels of trust, and thus mutual benefit, throughout 10 rounds of a game. When a BPD patient is the trustee, however, they take umbrage to any perceived slights and are unable to accurately detect attempts to repair a breakdown in trust, with the result that payoffs to both parties in the second half of the game are dramatically reduced. An fMRI study using the repeated trust game with BPD and healthy controls showed that activity in the anterior insula, a region known to be involved in a variety of social cognition tasks, corresponded with both the investment they received from their partner in the trust game and the amount of money they returned to their partner. In contrast, the BPD patients did not show a relationship between their insula function and either the amounts they received or the amounts they returned (King-Casas et al., 2008).

While the main body of the King-Casas et al. article focuses on the insula result, the online supplemental material notes bilateral activations in the fusiform gyrus and adjacent brain regions ((32, –64, –16, $k = 24$, $z = 4.01$) (–32, –76, –12, $k = 50$, $z = 4.55$) (32, –80, –12, $k = 23$, $z = 4.17$)). Of particular interest, this study provides the kind of minimal visual stimuli typical of a neuroeconomics experiment. The subject essentially just views bar graphs of the payoffs. Thus, we have no faces involved (the typical interpretation of fusiform activity in many social cognition experiments) and no obvious variability in the visual stimuli that would explain the pattern of activity in the fusiform and surrounding areas.

Further searches revealed a common pattern of fusiform activity in a number of articles involving norm violation and/or BPD patients where the activity was either unmentioned in the body of the article or given a single sentence talking about the role of the fusiform in face processing. When borderline patients viewed a set of images from the International Affective Picture System that contained scenes of violence and danger, crying children, or mutilated bodies, they had large bilateral responses in their fusiform (–40, –51, –17, $k = 418$, $t = 10.90$) (40, –51, –17, $k = 571$, $t = 11.62$) (Herpertz et al., 2001), a result that was largely replicated later (–38, –54, –18, $z = 4.15$, $k = 109$) (Koenigsberg et al., 2009). When the stimuli were positively valenced faces (–50, –62, –10; $t = 5.27$) or negatively valenced faces (–46, –70, –4, $t = 4.84$), the BPD patients had activations in areas overlapping the main finding in this article (Guitart-Masip et al., 2009). And, when the task was observing images triggering thoughts about failed social attachment, the fusiform was also implicated in BPD (42,–78,–12, $z = 4.26$) (Buchheim et al., 2008). All of this work suggests that the fusiform functions unusually in BPD patients, a population known to have tremendous difficulty with social norms.

The fusiform is also involved in studies where all the participants are nonclinical and the stimuli involved social norm violation. With verbal narration of intentional or unintentional norm violations, substantial activity in the fusiform is observed (–20, –82, –10, $z = 4.32$, $k = 686$) (–30, –80, –12, $z = 3.93$, $k = 444$) (Berthoz, Armony, Blair, & Dolan, 2002). Video of people being attacked with weapons also provokes fusiform activity (–46, –58, –22, $t = 5.11$, $k = 251$) (Nummenmaa, Hirvonen, Parkkola, & Hietanen, 2008). A variety of other norm violations such as people being deceived (46, –36, 20, $z = 3.27$, $K = 45$) (Grezes, Berthoz, & Passingham, 2006), having the threat of sanctions for norm violation in the trust game (36, –48, –16, $z = 3.63$, $K = 18$) (Li, Xiao, Houser, & Montague, 2009), being judged by others for violating social norms (no coordinates provided in the article) (Izuma, Saito, & Sadato, 2010), making a decision while under the threat of social sanction (–28, –68, –18, $z = 3.91$) (Spitzer, Fischbacher, Herrnberger, Gron, & Fehr, 2007), and deciding difficult moral dilemmas described with colorful language (27, –24, –30, $k = 52$, $F = 9.81$) (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006) all involve the fusiform. A number of these studies do not involve facial or other visual stimuli, and the fusiform results typically receive little or no mention in the articles.

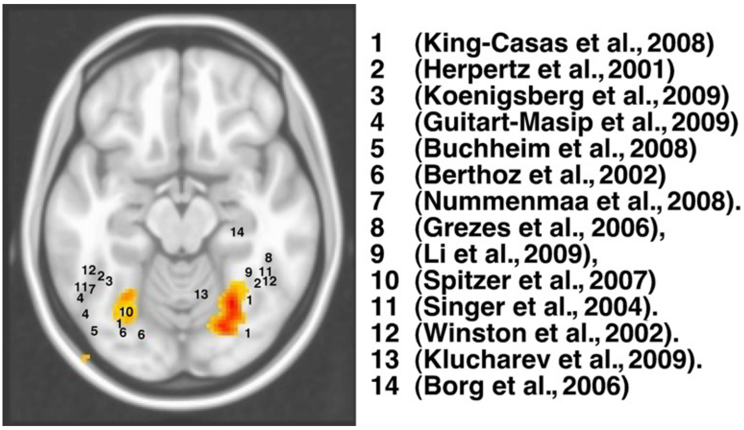

Figure 5. Coordinates of peak activations reported in 14 brain imaging studies of norm violation with borderline personality disorder patients and healthy participants (1–5) and healthy participants (6–14).

A few studies do provide discussions of the fusiform or adjacent activations in the context of social norms. Viewing the faces of intentional cooperators correlates with bilateral activation in the fusiform (42, –48, –33, $z = 3.70$) (–51, –63, –21, $z = 4.13$) (Singer, Kiebel, Winston, Dolan, & Frith, 2004). Viewing untrustworthy faces triggers bilateral activation of the fusiform (44, –46, –22; $z = 3.58$; –48, –48, –24; $z = 3.60$) (Winston, Strange, O'Doherty, & Dolan, 2002). And succumbing to pressure to conform with social norms is related to large deactivations encompassing the bilateral fusiform gyrus (10, –58, 4, $z = –5.61$, $K = 1{,}588$) (Klucharev, Hytonen, Rijpkema, Smidts, & Fernandez, 2009).

In Figure 5, we have overlaid the points of peak activation for fourteen studies involving norm violation in BPD and healthy participants (1–5) or only healthy participants (6–14) onto an axial slice of the brain with the results from our study for comparison. The spatial extent of many of these activations is quite large, with many of them being the largest activations in the study and extending into both hemispheres.

So why might the fusiform, a region typically considered to be involved in high-order visual processing, categorization, or face recognition be responsive to violations of social norms? A basic consequence of Charles Darwin's (1859/1996) ideas about evolution is that "an organ originally constructed for one purpose . . . may be converted into one for a wholly different purpose." Thus, some of the first fish to go on land, like tiktaalik, move about on fins that presage the legs of their tetrapod descendants.

This evolutionary principle is often found in the layering of social brain function onto regions where analogous, more basic cognitive processes are found. Thus, the social pain of exclusion activates a region in the anterior cingulate known to be responsive to the experience of physical pain (Eisenberger, Lieberman, & Williams, 2003), the revulsion to incest and iniquity engage a portion of the insula known to be involved in revulsion to infection (Borg, Lieberman, & Kiehl, 2008), and the neurons we use to control our own motor movement fire when we observe that movement in others (Iacoboni, 2008). A potential explanation then is that visual detection of socially relevant phenomena like recognition of faces and affect engaged the early analogues of the fusiform in our primate ancestors (Rajimehr, Young, & Tootell, 2009). As the complexity of the primate social world evolved, increasing levels of social cognition would demand additional ability to detect violations of social norms (Schreiber, 2007).

Processing visual evidence of social norm violation would likely require identification of the facial expressions and categorization of behaviors as cues to intentionality (de Waal, 1998) both when one violates a norm and when one observes others violating norms. If detection of visual evidence of norm violation increases functional connectivity with bottom-up and top-down processes, then it is easy to imagine the fusiform being engaged for norm violation even in contexts where the visual information is not paramount.

These results suggest support for Kurzban et al.'s (2001) contention that it is coalitional cognition, rather than race per se that is driving much of the observed difference in racial attitudes. When implicit stereotypes are defeated by the explicitly presented context, then measures of threat, dehumanization, and norm violation are reduced. These results suggest that norm-consistent images of African Americans are perceived as part of an "us," rather than a "them." Just as with Kurzban et al.'s (2001) results, the team membership trumps racial identity with minimal exposure.

Furthermore, the results in the amygdala have been traditionally interpreted as connected with implicit and automatic attitudes rather than with deliberate judgments. Such an interpretation would be consistent with Kurzban et al.'s (2001) findings, although it is important to note that there are many instances shown where deliberative processes down-regulate the automatic amygdala activations (e.g., Lieberman et al., 2005). It may be that the norm-consistent images evoke explicit mental processes, which alter the amygdala function. However, the fMRI data in the present study does not support such an interpretation. Further study would be needed to tease out such a theory.

Further explorations along the line of Kurzban et al.'s (2001) work would also be insightful. How would the brain regions discussed in this article respond to norm violation when ingroup membership has already been firmly established? Would individuals present evidence of the type of heightened response that comes when betrayal is unexpected because of prior loyalty? Would we be able to further disentangle norm violation from outgroup effects? And would we see the same kind of inability to erase the effecomplementcts of gender that Kurzban et al. (2001) identifies and contrasts with the erasable effects of race? On this last point, Kurzban et al. (2001) argues that evolutionary psychology suggests that gender is also salient for reproduction and long-term survival, whereas coalition membership is expected to be more permeable. Using similar reasoning, one might predict that serious norm violation consigns another to outgroup status (i.e., banishment), regardless of the strength of previous ingroup attachment.

Future work could test the implications of this article using large-*n* experimental surveys in which subjects are provided short vignettes where race and norm violation are varied. Evaluating priming effects for positive or negative emotions, for judgmental or laudatory adjectives, and for human or animal words could provide further evidence that automatic implicit associations are altered by the context of a racial encounter. Since the number of trials and participants limits this experiment, further replication would illuminate and solidify the findings. Using an event-related design in which images are shown individually could potentially trade some statistical power for an improved ability to identify the effects of particular images. And, replicating the study using a subset of images where faces are obscured would improve the ability to interpret the fusiform activity.

Conclusion

Functional brain imaging continues to develop as a new tool for exploring the mental processes that are involved in racial attitudes and social norms. Being able to get beyond behavioral data and self-reports allows us more direct evidence about how the brain is engaged in social cognition and what neural mechanisms might be active. The results in this article fit with the general trend towards seeing racial attitudes as part of a larger framework of ingroup/outgroup phenomena, but show the interplay of social norm violation as an important intervening force.

This article also points to two significant limitations in the existing functional imaging literature. First, while the use of disembodied faces in functional imaging studies is typical, the human brain did not evolve for seeing Cheshire–cat like grins devoid of context. These experiments have been useful in that they provide some evidence about the implicit associations and stereotypes that we carry about African Americans or other minority groups. However, implicit attitudes are not the only kind. The dual-process models of social psychology are important because they demonstrate the limits of introspection and show how much sway subconscious mental processes have. Nonetheless, our explicit perceptions and attitudes have been shown again and again to override the implicit ones. And furthermore, when we encounter people, it is usually heavily contextualized. The use of rich, natural stimuli like the images in this study should be a regular complement to the decontextualized stimuli that is more commonly presented.

Second, the functional brain imaging literature generates gigabytes of data with each experiment. This is orders-of-magnitude greater than previous generations of surveys and experiments. The downside of the cornucopia is that repeated patterns that do not fit previous expectations can lie dormant in the midst of that data. The role of the fusiform gyrus postulated in this article is supported by results in a number of functional imaging studies that do not even discuss the fusiform or consign it to tables, figures, or supplementary materials. The inference by analogy that predominates in functional imaging has path dependency as a byproduct. When a conventional interpretation of neural activity has become established, then the plethora of data makes it easy to instead focus on the easily comprehensible result. This happened in many of the articles cited here despite the fact that the fusiform results were often the largest and most statistically significant. However, because their substantive significance was unclear and an easy analogy was unavailable, they were often simply not discussed.

This problem is exacerbated by limitations that stem from the conventions about how fMRI data is reported. While tables reporting peak activations are useful, the spatial extent of the activations in these studies was sometimes so large that they overlapped areas far beyond the anatomy labeled in the tables. Only when articles also included visualizations of the data was it possible to see that the activity in these other studies overlapped or was adjacent to the results presented here. With a variety of naming conventions (e.g., Broadmann areas, Tailarach coordinates, MNI coordinates, and a host of different labels for the same anatomy), it is extremely difficult to search the existing literature for other relevant findings in a coherent manner. Attempts to remedy this include databases of reported coordinates of activations. However, in cases of broad spatial activity (as in many of the articles discussed above) even this would provide an inadequate solution. Since the National Institutes of Health has already coordinated a file format that all major functional imaging software uses, archiving the image files that are the result of a whole brain analysis should be relatively easy. As this article demonstrates, latent patterns likely exist among many results from studies that have already been done.

In this article, we have demonstrated that patterns of brain activity often attributed to racial stereotypes might be more appropriately ascribed to stereotypes rather than race per se. We have also identified an additional potential role for the fusiform gyrus in the perception of social norm violations. Further research will be required to confirm this possible connection. If so, we might have a new mechanism for investigating how the brain processes social phenomena and a deeper understanding of how dysfunctions emerge. If Borderline Personality Disorder does involve an abnormal function of the fusiform that is connected to the disease's dysregulation as regards social norms, then we may have also uncovered a mechanism that might lead to more effective treatment for a population that suffers substantially and involves large societal costs.

Understanding the role of social norms and ingroup/outgroup dynamics is at the core of political science. Developing a deeper comprehension of the neural mechanisms involved will alter our theories of human nature (Fowler & Schreiber, 2008) and thus alter the more macro theories that guide our design of political institutions. While functional imaging is not yet often used by political scientists, it has the potential to add to the many tools we already rely on.

Appendix

Subjects

For this study, we recruited 19 healthy undergraduate students in three groups. Six of the students were recruited from the college Republican Club, six were recruited from the college Democrat Club, and seven were undergraduates not affiliated with a campus political organization. All of the participants were screened with a 16-item political knowledge questionnaire. As expected, the political club members scored high on this measure with a mean political knowledge score of 15.2 ± 0.9. The unaffiliated undergraduates were solicited through email. We received 89 responses with a mean political knowledge score of 12.0 ± 2.9. In order to maximize the difference between the politically knowledgeable and the politically novice groups, we recruited seven undergraduates with the lowest political knowledge scores (mean 7.9 ± 1.9) to participate in the brain imaging portion of the experiment.

Because race perception was a central issue of our investigation, we initially included only the 18 Caucasian participants in the analysis of the brain imaging portion of the experiment. However, both a standard and robustness analysis of the fMRI data showed no significant differences from including the nineteenth participant who was of Asian descent. The 12 political club members (five females) had a mean age of 21.1 ± 2.1, whereas the seven political novices (six females) had a mean age of 20.2 ± 1.5. Participants gave informed consent following the guidelines of the college Institutional Review Board. All subjects were screened to rule out medication use, a history of neurological or psychiatric disorders, head trauma, substance abuse, or other serious medical conditions.

Stimuli

In order to enhance the external validity of the materials presented in this study, we collected images of African Americans and European Americans from a wide variety of magazines and websites. We selected images that were easily comprehendible, with visual cues that would allow subjects to grasp what they were seeing. We divided images into two distinct groupings. For the norm-consistent group, we found images of people that were consistent with mainstream cultural values. This group consisted of teachers, students, police officers, doctors, families, business people, and affectionate heterosexual couples. For the norm-violating group, we were looking for images that violated mainstream values. In this set we had images of gang members, rappers, punk rockers, rioters, homeless people, neo-Nazis, prison inmates, and mug shots of convicted sex offenders.

All of the images were either from digital sources or digitally scanned. They were all rendered in black and white to create consistency from the mix of color and black and white original source images. All of the images were resized to be similar when presented on the screen.

From a large pool of images, we selected 10 images for each of the groups in the two-by-two design (Race × Norm). As the brain imaging experiment called for a block design to maximize the likelihood of obtaining a strong BOLD fMRI response, we randomly sorted the 10 images of each cell of the design into two, five-image blocks. While in the scanner, subjects were presented with each image for four seconds, one after the other for a total of 20 seconds per block. All of the images in each block were of the same race and norms category. After each block of images, the subjects saw a blank screen for 20 seconds, followed by another block of images. In total, the subjects saw eight blocks of images over the course of the experiment. The subjects were asked to look at the images and told that they would be asked questions about what they saw.

Functional Imaging Data Acquisition and Analysis

Images were acquired using a GE 3.0T MRI scanner with an upgrade for echo-planar imaging (EPI) (Advanced NMR Systems, Inc.). A 2D spin-echo image (TR = 4,000 ms; TE = 40 ms, 256 by 256, 4-mm thick, 1-mm spacing) was acquired in the sagittal plane to allow prescription of the slices to be obtained in the remaining sequences. This sequence also ensured the absence of structural abnormalities in the brain of the enrolled subjects. For each subject, a high-resolution structural T2-weighted EPI volume (spin-echo, TR = 4,000 ms, TE = 54 ms, 128 by 128, 26 slices, 4-mm thick, 1-mm spacing) was acquired coplanar with the functional scans. One functional EPI scan (gradient-echo, TR = 4,000 ms, TE = 25 ms, flip angle = 90, 64 by 64, 26 slices, 4-mm thick, 1-mm spacing) was obtained for a duration of six minutes while the subjects viewed the images through MRI scanner-compatible video goggles. The functional scan covered the whole brain and was composed of 90 brain volumes. The order of the presentation of the blocks of images was randomized for each subject.

GE image files were converted into Analyze files and processed with FSL (http://www.fmrib.ox.ac.uk/fsl). Brain volumes within each fMRI run were motion-corrected using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002). Spatial smoothing was applied using a Gaussian-weighted kernel of 5 mm at full-width half-maximum, and data were high-pass filtered with sigma = 15.0 seconds and intensity normalized. Functional images were first registered to the coplanar high-resolution structural T2-weighted EPI volume after nonbrain structures had been removed from the coplanar high-resolution T2-weighted EPI volume using BET (Smith, 2002). Functional images were subsequently registered to the Montreal Neurological Institute Talairach-compatible MR atlas averaging 152 normal subjects using FLIRT (Jenkinson et al., 2002).

To analyze the data, we used two different statistical approaches. In the whole brain analysis we evaluated the time series data of the BOLD fMRI signal for each of the cubic millimeter voxels in the brain. This whole brain analysis allows for a broader set of hypothesis tests; for instance, looking for all brain regions that are more active when responding to norm-violating images than norm-consistent images. The other method we used is known as region of interest analysis (ROI). In this method, the data for all the voxels in a particular brain region are grouped and extracted for analysis.

Whole brain statistical analyses were carried out at two levels: the level of the individual and a higher-order intergroup comparison level. Time-series statistical analyses were carried out using FILM with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001). Higher-level statistics were carried out using mixed-effect (random effects) implemented in FLAME (FMRIB's Local Analysis of Mixed Effects) (Beckmann, Jenkinson, & Smith, 2003). *Z*-image statistics were performed with a threshold of $Z = 2.3$ at voxel level (Forman et al., 1995; Friston, Worsley, Frackowiak, & Mazziotta, 1994; Worsley, 2001; Worsley, Evans, Marret, & Neelin, 1992). After the initial analysis with 18 Caucasian subjects, all 19 subjects (18 Caucasians, 1 Asian) were included and the data reanalyzed using both the standard two-stage FLAME analysis and the robust analysis. The reanalysis indicated no significant differences, so the results of the robust analysis with 19 subjects were presented in the article.

ROI analysis was carried out using the Featquery program to extract the brain regions of interest. The ROIs for the amygdala were generated from the Harvard-Oxford Subcortical Structural Atlas. The ROI in the mPFC was generated using a 5-mm sphere centered at coordinates (–2, 48, −7) previously reported as being involved in mentalizing processes when viewing social outgroups (Freeman et al., 2010; Harris & Fiske, 2006).

ACKNOWLEDGMENTS

Correspondence concerning this article should be sent to Darren Schreiber, Department of Political Science, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093. E-mail: darren.schreiber@gmail.com

REFERENCES

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268–277.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in fMRI. *Neuroimage, 20*(2), 1052–1063.

Berman, M. G., Park, J., Gonzalez, R., Polk, T. A, Gehrke, A., Knaffla, S., & Jonides, J. (2010). Evaluating functional localizers: The case of the ffa. *Neuroimage, 50*(1), 56–71. doi:10.1016/j.neuroimage.2009.12.024

Berthoz, S., Armony, J. L., Blair, R. J., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain, 125*, 1696–1708.

Bobo, L. (1983). Whites' opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology, 45*, 1196–1210.

Borg., J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18*, 803–817.

Borg., J. S., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience, 20*, 1529–1546. doi:10.1162/jocn.2008.20109

Buchheim, A., Erk, S., George, C., Kachele, H., Kircher, T., Martius, P., & Walter, H. (2008). Neural correlates of attachment trauma in borderline personality disorder: A functional magnetic resonance imaging study. *Psychiatry Research, 163*, 223–235. doi:S0925–4927(07) 00145-X [pii] 10.1016/j.pscychresns.2007.07.001

Bush, G., & Shin, L. M. (2006). The multi-source interference task: An fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. *Nature Protocols, 1*, 308–313. doi:10.1038/nprot.2006.48

Cant, J., & Goodale, M. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex, 17*(3), 713–731. doi:10.1093/cercor/bhk022

Carver, C. S., Glass, D. C., & Katz, I. (1978). Favorable evaluations of blacks and the handicapped: Positive prejudice, unconscious denial, or social desirability? *Journal of Applied Social Psychology, 8*(2), 97–106.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248*(4962), 1556–1559.

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 pet and fMRI studies. *Brain Research Reviews, 58*(1), 57–70. doi:S0165-0173(07) 00248-2 [pii] 10.1016/j.brainresrev.2007.10.012

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806–813. doi:PSCI760 [pii] 10.1111/j.0956-7976.2004.00760.x

Darwin, C. (1996 [1859]). *The origin of species*. Oxford: Oxford University Press.

Davis, D., & Silver, B. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science, 47*, 33–45.

de Waal, F. B. M. (1998). *Chimpanzee politics: Power and sex among apes* (Rev. ed.). Baltimore: Johns Hopkins University Press.

Descartes, R. (1979). *Meditations on first philosophy: In which the existence of God and the distinction of the soul from the body are demonstrated*. Indianapolis, IN: Hackett Publishing Co. (Original work published 1641).

Devlin, J. T., Rushworth, M. F., & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia, 43*(1), 69–74. doi:S0028-3932(04) 00153-8 [pii] 10.1016/j.neuropsychologia.2004.06.013

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature, 427*(6972), 311–312. doi:10.1038/427311a 427311a

Eberhardt, J. L. (2005). Imaging race. *American Psychologist, 60*, 181–190. doi:2005-01817-004 [pii] 10.1037/0003-066X.60.2.181

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302*(5643), 290–292.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33*, 636–647.

Fowler, J. H., & Schreiber, D. (2008). Biology, politics, and the emerging science of human nature. *Science, 322*(5903), 912–914. doi:322/5903/912 [pii] 10.1126/science.1158188

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

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., & Mazziotta, J. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping, 1*, 214–220.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197.

Gilens, M. (1999). *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. Chicago: University of Chicago Press.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differntial responses in the fusiform region to same-race and other-race faces. *Nature, 4*(8), 845–850.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41. doi:2009-08950-006 [pii] 10.1037/a0015575

Grezes, J., Berthoz, S., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *Neuroimage, 30*, 601–608. doi:S1053-8119(05) 00751-2 [pii] 10.1016/j.neuroimage.2005.09.038

Guitart-Masip, M., Pascual, J. C., Carmona, S., Hoekzema, E., Berge, D., Perez, V., & Vilarroya, O. (2009). Neural correlates of impaired emotional discrimination in borderline personality disorder: An fMRI study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 33*, 1537–1545. doi:S0278-5846(09) 00293-0 [pii] 10.1016/j.pnpbp.2009.08.022

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Review Neuroscience, 2*, 685–694.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science, 17*, 847–853. doi:PSCI1793 [pii] 10.1111/j.1467-9280.2006.01793.x

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *Neuroreport, 11*, 2351–2355.

Hatchett, S., & Schuman, H. (1975). White respondents and race-of-interviewer effects. *Public Opinion Quarterly, 39*, 523–528. doi:10.1086/268249

Herpertz, S. C., Dietrich, T. M., Wenning, B., Krings, T., Erberich, S. G., Willmes, K., & Sass, H. (2001). Evidence of abnormal amygdala functioning in borderline personality disorder: A functional MRI study. *Biological Psychiatry, 50*, 292–298. doi:S0006-3223(01) 01075-7

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*, 575–604. doi:10.1146/annurev.psych.53.100901.135109

Huddy, L., & Feldman, S. (2009). On assessing the political effects of racial prejudice. *Annual Review of Political Science, 12*, 423–447.

Iacoboni, M. (2006). Failure to deactivate in autism: The co-constitution of self and other. *Trends in Cognitive Science, 10*, 431–433.

Iacoboni, M. O. (2008). *Mirroring people: The new science of how we connect with others*. New York: Farrar, Straus, and Giroux.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Science, 13*, 524–531. doi:S1364-6613(09) 00216-2 [pii] 10.1016/j.tics.2009.10.002

Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience, 5*, 133–147. doi:914151038 [pii] 10.1080/17470910903202559

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*, 825–841.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*(5386), 108–111.

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science, 323*(5911), 276–278. doi:323/5911/276 [pii] 10.1126/science.1164951

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science, 321*(5890), 806–810. doi:321/5890/806 [pii] 10.1126/science.1156902

Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A., & Fernandez, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61*(1), 140–151. doi:S0896-6273(08) 01020-9 [pii] 10.1016/j.neuron.2008.11.027

Koenigsberg, H. W., Siever, L. J., Lee, H., Pizzarello, S., New, A. S., Goodman, M., & Prohovnik, I. (2009). Neural correlates of emotion processing in borderline personality disorder. *Psychiatry Research, 172*(3), 192–199. doi:S0925-4927(08) 00096-6 [pii] 10.1016/j.pscychresns.2008.07.010

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

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Science USA, 98*(26), 15387–15392.

Li, J., Xiao, E., Houser, D., & Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 16835–16840. doi:0908855106 [pii] 10.1073/pnas.0908855106

Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *Lancet, 364*(9432), 453–461. doi:10.1016/S0140-6736(04) 16770-6

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

Lieberman, M., Schreiber, D., & Ochsner, K. (2003). Is political sophistication like learning to ride a bicycle? How cognitive neuroscience can inform research on political thinking. *Political Psychology, 24*, 681–704.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 11832–11835.

Minshew, N. J., & Keller, T. A. (2010). The nature of brain dysfunction in autism: Functional brain imaging studies. *Current Opinions in Neurology, 23*(2), 124–130. doi:10.1097/WCO.0b013e32833782d4

Northoff, G., Qin, P., & Nakao, T. (2010). Rest-stimulus interaction in the brain: A review. *Trends in Neurosciences, 33*(6), 277–284. doi:10.1016/j.tins.2010.02.006

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage, 43*, 571–580. doi:S1053-8119(08) 00929-4 [pii] 10.1016/j.neuroimage.2008.08.014

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Reviews of Psychology, 60*, 339–367. doi:10.1146/annurev.psych.60.110707.163607

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 11458–11463. doi:10.1073/pnas.172403899 172403899

Phelps, E. A., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia, 41*(2), 203–208. doi:S0028393202001501

Phelps, E. A., O'Connor, K. J., Cunningham, W., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Mahzarin, R. B. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729–738.

Pujol, J., Harrison, B. J., Ortiz, H., Deus, J., Soriano-Mas, C., Lopez-Sola, M., & Cardoner, N. (2009). Influence of the fusiform gyrus on amygdala response to emotional faces in the non-clinical range of social anxiety. *Psychological Medicine, 39*, 1177–1187. doi:S003329170800500X [pii] 10.1017/S003329170800500x

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 676–682.

Rajimehr, R., Young., J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 1995–2000. doi:10.1073/pnas.0807304106

Reddy, L., Moradi, F., & Koch, C. (2007). Top-down biases win against focal attention in the fusiform face area. *Neuroimage, 38*, 730–739. doi:10.1016/j.neuroimage.2007.08.006

Rilling, J. K., Dagenais, J. E., Goldsmith, D. R., Glenn, A. L., & Pagnoni, G. (2008). Social cognitive neural networks during in-group and out-group interactions. *Neuroimage, 41*, 1447–1461. doi:10.1016/j.neuroimage.2008.03.044

Scholz, J., Klein, M. C., Behrens, T. E., & Johansen-Berg., H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience,* 1370–1371. doi:nn.2412 [pii] 10.1038/nn.2412

Schreiber, D. (2007). Political cognition as social cognition: Are we all political sophisticates? In W. R. Neuman, G. E. Marcus, A. N. Crigler, & M. MacKuen (Eds.), *The affect effect: Dynamics of emotion in political thinking and behavior* (pp. 48–70). Chicago: University of Chicago Press.

Schreiber, D. (2011). From SCAN to neuropolitics. In P. K. Hatemi & R. McDermott (Eds.), *Man is by nature a political animal*. Chicago: University of Chicago Press.

Sears, D., Van Laar, C., Carrillo, M., & Kosterman, R. (1997). Is it really racism? The origins of white Americans' opposition to race-targeted policies. *Public Opinion Quarterly, 61*, 16–53.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Review, 32*, 811–830. doi:S0149-7634(08) 00007-9 [pii] 10.1016/j.neubiorev.2007.12.002

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*, 653–662. doi:S0896627304000145

Smith, C. D., Lori, N. F., Akbudak, E., Sorar, E., Gultepe, E., Shimony, J. S., & Conturo, T. E. (2009). MRI diffusion tensor tracking of a new amygdalo-fusiform and hippocampo-fusiform pathway system in humans. *Journal of Magnetic Resonance Imaging, 29*, 1248–1261. doi:10.1002/jMRI.21692

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*, 143–155.

Sniderman, P. M., & Carmines, E. G. (1997). *Reaching beyond race*. Cambridge, MA: Harvard University Press.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*, 185–196. doi:S0896-6273(07) 00709-X [pii] 10.1016/j.neuron.2007.09.011

Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin, 35*, 321–335. doi:0146167208327743 [pii] 10.1177/0146167208327743

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science, 19*, 1131–1139. doi:PSCI2214 [pii] 10.1111/j.1467-9280.2008.02214.x

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*, 829–858. doi:10.1002/hbm.20547

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*, 829–841.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science, 16*, 56–63. doi:PSCI780 [pii] 10.1111/j.0956-7976.2005.00780.x

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*, 277–283. doi:10.1038/nn816 nn816

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology, 79*, 1574–1578.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage, 14*, 1370–1386.

Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: An introduction to methods*. Oxford: Oxford University Press.

Worsley, K. J., Evans, A. C., Marret, S., & Neelin, P. (1992). A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism, 12*, 900–918.

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