Pictures in our heads:

Contributions of fMRI to the study of prejudice and stereotyping

David M. Amodio and Matthew D. Lieberman

New York University and University of California, Los Angeles

To Appear in T. Nelson (ed.) *Handbook of Prejudice, Stereotyping, and Discrimination*.

Earlbaum Press.

Corresponding Author:
David Amodio
david.amodio@nyu.edu

In 1922, Walter Lippman famously referred to stereotypes as "pictures in our heads." His comment presaged nearly a century of research on the how perceptions of stigmatized social groups are represented in the mind. In this chapter, we describe how the most recent addition to the prejudice researcher's methodological toolbox – functional magnetic resonance imaging (fMRI) – allows researchers to measure patterns of neural activity associated with prejudice, stereotyping, and discrimination (Figure 1). fMRI is a technique for measuring changes in blood flow in the brain. As neurons in the brain fire, their energy is depleted. Tiny capillaries throughout the brain deliver oxygenated blood supplies to replenish neuron's energy stores. Oxygenated blood contains more ionized hemoglobin molecules, and changes in blood oxygen-dependent (BOLD) signal can be detected using magnetic resonance technology (see Huettel, Song, & McCarthy, 2004, for an in-depth description of fMRI methodology). The assumption in fMRI research is that increases in blood flow to a particular region of the brain is associated with a greater degree of neuronal activity in the preceding seconds. When placed in the hands of prejudice researchers, fMRI provides a way to study Lippman's "pictures in our heads" by examining patterns of activity in our brains (i.e., pictures *of* in our heads).

In this chapter, we describe how neuroimaging methods have been used to study different components of racial bias, and how this research has contributed to theoretical advances in the field of intergroup bias. A second goal of this chapter is to present the extant findings on prejudice and stereotyping in a framework that emphasizes their role in the regulation of behavior. We begin with a brief review of the social cognition literature on prejudice and stereotyping to provide context for the body of recent fMRI studies in this area.

Social cognition research on prejudice and stereotyping

In his book, *The Nature of Prejudice*, Allport (1954) observed that when it comes to race relations, many White Americans live in a state of conflict: on one hand, they may be ideologically opposed to prejudice, but on the other, they possess underlying tendencies to think and act in racially biased ways. More recent conceptualizations of Allport's "state of conflict" suggest that people may hold explicit egalitarian beliefs while simultaneously possessing implicit racial associations that operate automatically in subconscious mental processes (e.g., Devine, 1989; see also Wilson, Lindsey, & Schooler, 2000). The interplay of implicit associations and explicit beliefs has captured the attention of social cognition researchers in recent years, as reviewed in more detail elsewhere in this volume, and the majority of fMRI investigations of racial bias have been designed to address central issues in the social cognition of prejudice. To set the stage for our review, we begin with a brief review of the key socio-cognitive mechanisms of prejudice that have been of particular interest to researchers taking a neuroscience approach.

Automaticity of bias. Automatic forms of race bias are typically described in a few different ways: as an instantaneous gut-level feeling about a person or group, or alternatively as thought that spontaneously pops into one's head when upon encountering a member of a stigmatized social group (Fiske, 1998; Greenwald & Banaji, 1995). Still others have focused on motor components of automaticity, such as spontaneously activated behaviors that are engaged when exposed to an outgroup member (Bargh, Chen, & Burrows, 1996; see also James, 1890). Although these different forms of implicit bias have been documented, very little research has distinguished them at a theoretical level, and thus the assumption has been they are learned, expressed, unlearned, and regulated through the same set of mechanisms (but see Amodio & Devine, 2006).

The proposition that automatic forms of racial bias could be dissociated from consciously-held attitudes and beliefs was first demonstrated by Devine (1989). On the basis of research in cognitive psychology, Devine theorized that stereotypes were cognitive associations that could be over-learned through repeated exposure in one's cultural environment, such that they may be automatically activated in response to relevant stimuli (Meyer & Schvaneveldt, 1971; Shiffrin & Schneider, 1977). Her research showed that subconscious exposure to race-related words activated these stereotype constructs in subjects' mental representations, which in turn biased participants' impressions of novel individuals in stereotype-consistent ways. What was perhaps most interesting about her findings was that the automatic effects of stereotypes on behavior were not moderated by participants' level of explicit prejudice when they were unaware of the racial primes. Although implicit racial associated had been demonstrated in earlier work (Dovidio, Evans, & Tyler, 1986; Gaertner & McLaughlin, 1983), Devine (1989) suggested that Allport's (1954) "state of conflict" referred to a conflict between explicit beliefs and implicit stereotype associations.

Subsequent work focused on the automatic activation of negative evaluations of racial outgroups (i.e., implicit prejudices), such as in White people's responses to African Americans (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995). Whereas Devine (1989) examined the effects of subliminally-primed stereotypes on social judgments, most of the work investigating evaluative effects of bias have focused on the relationship between reaction-time—based measures of bias with outcomes such as social behavior and self-reported attitudes. Implicit prejudice is typically indicated by greater facilitation of responses to negative words or objects following exposure to Black faces than White faces (and relative to responses to positive stimuli). Research using reaction-time

measures has shown that implicit evaluations of Black people are generally unrelated to individuals' explicit attitudes, yet they predict biased patterns of nonverbal behaviors in actual and anticipated interracial interactions (Amodio & Devine, 2006; Dovidio, Kawakami, & Gaertner, 2002; Dovidio et al., 1997; Fazio et al., 1995; Henderson-King & Nisbett, 1996; McConnell & Leibold, 2001).

It is generally believed that implicit racial biases reflect exposure to biased patterns of racial associations in one's cultural milieu. Although little research has made direct connections between implicit bias and specific learning mechanisms (but see Olson & Fazio, 2006; Rydell & McConnell, in press), research suggests that implicit racial biases are learned passively, without one's deliberative intention to learn (Gregg, Seibt, & Banaji, 2006; Rydell & McConnell, in press). Indeed, much research examining associations between implicit and explicit racial responses have generally found modest relations (Blair, 2002), supporting the idea that automatic and consciously-held attitudes and beliefs arise from independent processes (Devine, 1989; Wilson et al., 2000). As a result, implicit and explicit biases have been shown to predict different forms of discrimination (Dovidio et al., 1997; Fazio et al., 1995). Yet the underlying mechanisms for how different forms of bias affect different types of behaviors remain poorly understood, in part because different underlying forms of implicit bias are difficult to parse using behavioral measures.

Regulating intergroup responses. How are automatic biases controlled? For egalitarians – those who reject prejudiced ideology – intentional intergroup behavior requires the regulation of unwanted automatic biases (Devine, 1989). Regulation is accomplished through controlled processing: the effortful and deliberative implementation of an intended response that overrides the influences of unwanted automatic biases, such as implicit prejudices and stereotypes (Shiffrin

& Schneider, 1977). Thus, egalitarians are expected to engage controlled processes in interracial interactions whereas racists would not. Although numerous studies have demonstrated the effectiveness of controlled processing in regulating intergroup responses, the social psychology literature lacks a mechanistic model for how controlled processes accomplish intentional responses in the face of automatic biases (D. T. Gilbert, Fiske, & Lindzey, 1998). How do controlled processes interface with behavior? What is being controlled – race-biased thoughts? Emotions? Behaviors? Are there multiple components of control? These important questions have been difficult to address using the tradition tools and theoretical models of social psychology, yet they are critical to our understanding of prejudice control, and of self-regulation more broadly.

An fMRI approach to the activation and regulation of intergroup responses

Over the past 15 years, a large body of accumulated findings attests to the power and pervasiveness of implicit racial biases as well as to human's great capacity to regulate their effects on behavior (Blair, 2001). Interestingly, this body of research is largely descriptive. There have been several demonstrations of implicit biases and efforts to control one's racial responses. But there hasn't been a clear, concrete theoretical explanation of what implicit bias is, what mechanisms facilitate its expression in behavior, and what mechanisms inhibit its expression. Without a strong theoretical model, efforts to predict the behavioral effects of implicit bias and to develop effective strategies for reducing implicit bias are limited. A major goal of the neuroscience approach to these enduring questions is to provide some theoretical scaffolding upon which further advances in the understanding of intergroup behavior may be built. Our review of the neuroimaging literature on prejudice and stereotyping begins by highlighting the contributions of fMRI research to the central social cognitive mechanisms of racial bias outlined

above (Table 1). We then describe some new directions in person perception that are relevant to issues of prejudice suggested by recent neuroimaging studies.

Neural mechanisms of implicit prejudice

Some of the earliest mergers of social psychological and cognitive neuroscience approaches were aimed at identifying the neural underpinnings of implicit prejudice (for review, see Lieberman, 2007). Behavioral neuroscience investigations of classical conditioning in rodents had identified the amygdala – a small set of nuclei located bilaterally in the medial temporal lobes – as critical for fear conditioning (Figure 2; Davis, Hitchcock, & Rosen, 1987; but see Davis & Whalen, 2001; Fendt & Fanselow, 1999; LeDoux, 1992). When describing research on the amygdala, it is important to note that interpretation of amygdala function have evolved considerably over the years, and although research continues to refine our understanding, functional explanations of the amygdala (as with most other brain structures) will likely undergo further revisions.

Early investigations of human amygdala function focused on the role of the amygdala in emotional processing, particularly as it pertains to the learning, perception, and expression of fear (Adolphs, Tranel, Damasio, & Damasio, 1995). Similarly, early neuroimaging studies found that presentations of fearful faces enhanced participants' amygdala activity, relative to neutral or happy facial expressions (Breiter, Rauch, Kwong, Baker, & et al., 1996; Morris, Frith, Perrett, Rowland, & et al., 1996). Later refinements to this body of work suggested that the amygdala serves as a low-level threat detector that is activated in response to stimuli that are potentially dangerous. Thus it was associated not just with fear, but also ambiguity, vigilance, arousal, and even uncertainly associated with positive outcomes (Whalen, 1998). Accumulating evidence continues to suggest that the amygdala responds to the emotional intensity of a stimulus (i.e., the

arousal component of affect) rather than to the valence of a stimulus (Anderson et al., 2003; Cunningham, Raye, & Johnson, 2004), although intensity tends to be greater for negative stimuli on average (Lang, Bradley, & Cuthbert, 1990; Cacioppo, Gardner, & Berntson, 1999).

Despite changes in functional interpretations of amygdala response, neuropsychological and neuroimaging research has consistently demonstrated that the amygdala operates at an automatic and unconscious level of processing. A seminal study by Bechara et al. (1995) examined the ability of patients with bilateral amygdala damage to learn in a classical-conditioning paradigm. In the task, participants viewed a series of colored shapes, some of which were paired with an aversive noise (a 100 dB blast of a boat horn). The researchers assessed learning in two ways, designed to test participants implicit vs. explicit processing. To assess explicit learning, participants were simply asked to report which stimulus was paired with the horn blast. To assess implicit learning, the researchers examined changes in participants' skin conductance levels when the condition stimulus appeared. Skin conductance reflects activity of the autonomic nervous system, and levels typically rise in anticipation of an aversive event. It was found that although the amygdala patients could correctly report the conditioned stimulus, they did not show the typical anticipatory rise in skin conductance, suggesting that the amygdala was important for implicit but not explicit processing. By contrast, a comparison group of patients with bilateral hippocampus damage were unable to report the conditioning contingency, yet their skin conductance levels displayed normal patterns of anticipatory autonomic responses when conditioned stimuli appeared, relative to stimuli that were not paired with the horn blast. Subsequent neuroimaging research showing that subliminal presentation of angry faces, masked by neutral faces, selectively activated the amygdala (Whalen et al., 1998), corroborating the notion that the amygdala operates at implicit level of processing.

To prejudice researchers, the amygdala seemed like an excellent candidate for a neural substrate of implicit prejudice. Research in social psychology has long suggested that feelings of fear may underlie implicit or gut-level negative evaluations of African Americans (Mackie & Smith, 1998; Smith, 1993), and so the amygdala seemed like an obvious choice. The first fMRI studies of prejudice measured brain activity while participants passively viewed faces of Black and White individuals. For example, Phelps et al. (2000) examined White American subjects' neural responses to unfamiliar Black faces, in comparison with White faces. Although the authors did not observe a significant increase in amygdala activity to Black vs. White subjects, there was a trend toward this effect. In addition, they showed that the degree of difference in amygdala activity to Black vs. White faces was correlated with participants' scores on a behavioral measure of implicit prejudice (the Implicit Associations Test, or IAT, Greenwald, McGhee, & Schwartz, 1998), as well as on a measure of startle-eyeblink response to Black vs. White faces that is known to be modulated by amygdala (Lang, Bradley, & Cuthbert, 1990; see also Amodio, Harmon-Jones, & Devine, 2003). This pattern of correlations provided the first evidence that amygdala activity might underlie implicit prejudice.

In the same year of Phelps et al.'s seminal paper, Hart et al. (2000) published research examining White and African American subjects' neural responses to faces of Black and White individuals. Hart et al. (2000) assessed neural activity to ingroup and outgroup faces in two blocks of trials (i.e., runs). Although amygdala activity to ingroup vs. outgroup faces did not differ during the first block of trials, a difference emerged in the second block, such that amygdala responses to ingroup faces were lower than responses to outgroup faces. The authors' interpretation of this effect was that in the first block, all faces were unfamiliar to participants, and the amygdala was similarly active to the ingroup and outgroup. However, by the second

block, participants had habituated to the ingroup faces, but not the outgroup faces. These effects were conceptually consistent with the findings of Phelps et al. (2000), in that they implicated the amygdala in implicit responses to race.

Significant differences in amygdala response to Black compared with White faces were initially reported by Amodio et al. (2003), who used the startle-eyeblink method to infer the degree of amygdala activation, and this pattern has since been replicated several times in fMRI studies using a range of experimental tasks (Cunningham, Johnson et al., 2004; Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005; Wheeler & Fiske, 2005). The strongest evidence to date that the amygdala may be involved in implicit prejudice was provided by Cunningham et al. (2004), in which participants were exposed to 30-msec presentations of Black and White faces (i.e., subliminal), masked by various shapes. Participants' task was to indicate whether the shape appeared on the left or right side of the screen. The authors found that subliminal presentations of Black faces elicited greater amygdala activity than White faces, and that the degree of increased amygdala activity to Black (vs. White) faces was associated with more anti-Black responses on an IAT assessing evaluative associations with Black vs. White faces. Wheeler and Fiske (2005) also observed greater amygdala activity in response to Black vs. White faces, but only when subjects' task was to categorize faces according to race. When the participant's task was to make an individuating inference from the face picture (e.g., guessing whether the target likes various vegetables) or when the task drew attention away from facial features of the target (e.g., when judging whether a small white dot was present in the picture), race effects for amygdala activity were not observed. On the surface, Wheeler and Fiske's (2005) finding that mere exposure to the faces did not activate the amygdala may appear to contradict the amygdala effects for subliminal pictures of Black faces observed by Cunningham et al.

(2004). However, we speculate that the lack of amygdala activity during the dot-finding task and individuation task in the Wheeler and Fiske (2005) study may have been related to a redirection of attentional resources associated with task demands. By contrast, the task used by Cunningham et al. (2004) was less difficult, and although participants were not aware of having viewed a face, ample attentional resources were available for subconscious processing of faces.

Although most research examining the amygdala as a substrate of implicit prejudice has focused on White American subjects, some theories of implicit race bias suggest that implicit prejudice is in part a cultural phenomena learned by all members of the culture, regardless of their race (Devine, 1989; Greenwald & Banaji, 1995; Rudman, 2004). If this is true, and then African American subjects should also show greater amygdala activity toward Black faces than White faces, despite the obvious fact that they rarely (if ever) hold explicit anti-Black prejudices. In line with this prediction, Lieberman et al. (2005) found that exposure to Black vs. White faces elicited greater amygdala activity among both White and African American participants. This finding is consistent with some behavioral research indicating anti-Black bias among African American subjects (Correll, Park, Judd, & Wittenbrink, 2002).

Richeson et al. (2003) examined White Americans' neural responses to images of Black vs. White faces and compared these responses with subjects' scores on an IAT measure of racial evaluations. Interestingly, the authors did not find the typical pattern of enhanced amygdala activity in response to Black vs. White faces, nor was change in amygdala activity associated with scores on the IAT. By contrast, regions of the prefrontal cortex (PFC) that are typically associated with executive function and working memory were more highly activated to Black than White faces and were positively correlated with implicit prejudice scores on the IAT. The authors interpreted this finding as reflecting participants' spontaneous attempt to control any

prejudiced thoughts that may have been caused by the pictures, and suggested that individuals with strong implicit prejudice may have been more likely to engage in such attempts. In summary, implicit prejudice has been the most well-studied component of intergroup bias in the fMRI literature. Across several studies using fMRI greater amygdala activation has been observed while White subjects viewed Black faces compared with White faces. Importantly, the interpretation that the difference in amygdala activity is associated with implicit prejudice has been validated in several studies through comparisons with behavioral and physiological assessments of implicit bias (e.g., Cunningham, Johnson et al., 2004; Phelps et al., 2000) and by comparing patterns of amygdala activation with known individual differences associated with implicit bias (Amodio et al., 2003).

Neural correlates of implicit stereotyping

Much research has focused on the role of the amygdala in evaluative and affective forms of implicit bias. But what about implicit stereotyping? Little, if any, research has yet explored these topics. However, recent theorizing by Amodio and Devine (2006) noted that implicit stereotyping relies upon representations of conceptual knowledge and associations, which are supported by neurocognitive systems for implicit semantic memory (also referred to as conceptual priming; Gabrieli, 1998). According to neuroscientific models of memory systems (e.g., Squire & Zola, 1996), semantic memory processes are generally supported by regions of the neocortex and not the regions of subcortex associated with implicit prejudice. Results from neuroimaging research on semantic memory and conceptual priming are somewhat mixed, yet an emerging pattern of findings suggests that conceptual priming involves regions of lateral temporal lobe (lTL) and ventral lateral PFC (Rissman, Eliassen, & Blumstein, 2003; Wible et al., 2006; Wig, Grafton, Demos, & Kelley, 2005; see Figure 3). On the basis of this body of

research, Amodio and Devine (2006) suggested that the mechanisms underlying implicit prejudice and implicit stereotyping are independent and dissociable, and are thus likely to be learned, expressed, regulated, and unlearned in somewhat different ways.

Research by Potanina, Pfeifer, Lieberman, and Amodio (2006) directly tested the hypothesis that implicit stereotyping should be uniquely associated with neural activity in the lTL and PFC (but not the amygdala), whereas implicit prejudice should be uniquely associated with activity in the amygdala (but not lTL or PFC). The task used by Potanina et al. (2006) was designed to engage participants in judgments of Black and White targets that relied on either basic affective or stereotypic information. The study was described as examining one's ability to infer information about a target person based on a picture of the person's face. In particular, participants were told that the study was testing whether they could accurately infer a person's preferences for certain activities, such as sports, or the likelihood that the target individual is the type of person the participant would be friends with. To strengthen the cover story and to lead participants to believe that we could later assess the accuracy of their judgments, participants first filled out questionnaires assessing their personal preferences for various activities/hobbies and for qualities they preferred in a friend. They were then told they would make judgments of pictures of people who had reported the same information (on friendship and activity preferences), such that we could check the accuracy of their inferences. Next, participants learned they would view pairs of people's faces and decide which of the pair was more likely to (a) be someone they would likely befriend (an affect-based judgment) or (b) preferred to engage in athletic activities (a more cognitive/stereotype-based judgment). Athletics was chosen because it is a central African American stereotype that is positive in valence and thus unlikely to involve negative affective processes (Devine & Elliot, 1995). Furthermore, the pair of faces

presented on each trial was always of the same race (Black, White, or Asian), and therefore judgments could not be influenced by participants' concerns about responding with prejudice. That is, issues of prejudice control were irrelevant when judging which of two Black individuals is more likely to be athletic or more likely to be friendly.

Consistent with their hypotheses, Potanina et al. (2006) observed greater activity in the in amygdala when participants judged Black face pairs on the basis of potential friendship, compared with White face pairs. Regions of neocortex associated with semantic processing were not observed for this contrast. On the other hand, the authors observed greater activity in the region of the left lateral temporal lobe and left PFC when participants judged Black vs. White face pairs on the basis of athleticism. However, this comparison did not elicit amygdala activity. These results provide the first evidence that distinct neural mechanisms appear to be associated with implicit prejudice and implicit stereotyping, as suggested by the cognitive neuroscience literature on memory. With evidence that difference memory systems underlie implicit prejudice and stereotyping, future research will be able to apply behavioral neuroscience models of learning to further our understanding of how implicit racial bias is learned and unlearned.

Neurocognitive mechanisms of control

Humans have a unique capacity for regulating their behaviors in order to behave in line with one's intentions. Understanding the way in which the mind carries out the process of self-regulation is a central concern among prejudice researchers. Social neuroscientists' research on this issue has largely followed from the broader cognitive neuroscience literature on control. One influential theory from this literature is that successful control involves the concerted activity of two independent processes for a) determining when control is needed and b) implementing the desired behavior despite unwanted tendencies (Botvinick, Braver, Barch, Carter, & Cohen,

2001). This model is built on the assumption that representations of response tendencies (e.g., motor plans) are spontaneously activated in the brain. Occasionally, two or more representations with conflicting response implications are activated at the same time and create the potential for unintended behavior. Botvinick et al. (2001) proposed a solution to crosstalk dilemma, whereby the degree of conflict in the system at any moment is represented in a *conflict monitoring* processes. In a sense, activity of the conflict monitoring component serves as a barometer of response conflict. As the level of conflict rises, the conflict-monitoring mechanism signals a second processes referred to as the *regulative* component for top-down control. The regulative process is responsible for intervening in crosstalk and deciding which of the competing responses should be implemented. This model is unique because it posits a bottom-up process for detecting the need for control, thereby dispensing with the "homunculus" idea assumed by most social-cognitive models in which a "little man" inside our heads "just knows" when to engage control. An important feature of Botvinick et al.'s (2001) model of control is that the two components – conflict monitoring and regulation – are associated with distinct neural substrates. Across several fMRI and PET studies, conflict monitoring has been associated with activity in the dorsal anterior cingulate cortex (dACC), a region of cortex that is proximal to the supplementary motor cortex and has strong connections to a wide range of neural structures (Figure 4). The regulative mechanism has been associated with the lateral prefrontal cortex (lPFC), a region previously associated with executive control and working memory functions (see Figure 3; S. J. Gilbert et al., 2006).

Botvinick et al's (2001) model of control has been very influential to researchers interested in the neural mechanisms of prejudice control. It is widely assumed that the process of regulating intergroup responses involves general mechanisms of control (as opposed to

specialized neural mechanisms for controlling racial biases). The role of the ACC as a conflict monitoring mechanism in the context of racial prejudice was first demonstrated using ERPs (Amodio et al., 2004). Amodio et al. showed that ERP responses arising from the dACC were larger on trials that activated automatic stereotypes that conflicted with participants' intended response (see also Amodio, Devine, & Harmon-Jones, under review; Amodio, Kubota, Harmon-Jones, & Devine, 2006), Using ERPs, Amodio et al. (2004) observed an increase in dACC activity when a response required inhibition 100 msec before the response was made. Moreover, participants showing greater sensitivity of this conflict system on error trials were better at inhibiting stereotypes throughout the task. However, although ERP measures permit researchers to examine patterns of neural firing as it changes over the course of milliseconds, certain neuroanatomical factors render ERPs more sensitive to activity in some brain regions than others. ERPs tend to be very sensitive to changes in the dACC, but not very sensitive to changes in areas of the lPFC that are important for controlled processing. For this reason, fMRI has been a more useful tool for studying the regulative component of control.

fMRI studies of prejudice control

fMRI provides much higher spatial resolution and coverage of frontal cortical processes than ERP measures, and therefore is a particularly useful tool for studying the control of prejudice. One of the first fMRI studies examining the control of prejudice was conducted by Cunningham et al. (2004). In their study, participants viewed faces of Black and White individuals and pictures of shapes. Their task was simply to indicate whether the stimulus appeared on the left or right side of their visual field. The authors observed greater amygdala activity to Black vs. White faces when faces were presented subliminally (i.e., for 30 msec), as described above. In contrast, when faces were presented for 525 msec and thus consciously

perceived, activity in the dACC and lPFC – regions implicated in Botvinick et al.'s (2001) control network – were stronger in response to Black vs. White faces. These results replicated the findings of Richeson et al. (2003), in which passive viewing of Black vs. White faces elicited ACC and PFC activity, and suggest that some element of control was more active among participants as they viewed Black faces. In addition, Cunningham et al. (2004) observed activity in the ventral region of lPFC. Whereas dorsal regions of lPFC have been primarily implicated in the implementation of an intended response, some theorizing suggests that the ventral lPFC may be involved in the inhibition of an unwanted behavioral or emotional response (Aron, Robbins, & Poldrack, 2004; Lieberman et al., in press; Ochsner, Bunge, Gross, & Gabrieli, 2002). Cunningham et al.'s (2004) results suggest that both forms of control may be involved when regulating prejudice.

The findings that viewing Black compared with White faces elicits activity in frontal cortical regions implicated in control raise important questions regarding the nature of "control" in the context of experimental studies. That is, what exactly is being controlled? Given that these activations were observed when participants either viewed faces passively or simply decided which side the screen the stimulus appeared, it is not clear whether these activations were associated with the intentional modulation of a thought, feeling, or behavior related specifically to responding without prejudice.

In an effort to begin to address some of the ambiguities of the fMRI literature on prejudice control, Amodio & Potanina (2006) recently used fMRI to examine subjects' neural activity while they made decisions directly that could be influenced by explicit motivations to respond without prejudice. The authors used the same paradigm as Potanina et al. (2006) described above, in which faces were judged according to the likelihood of friendship or

athleticism, except that participants made decisions about mixed-race (i.e., Black vs. White) face pairs in addition to same-race pairs. When making judgments about mixed-race pairs, we expected that subjects' concerns about appearing biased would become relevant, and thus they would be more deliberative in their ratings and try to respond in a way that did not reveal bias. In line with our hypotheses, we found that judgments of mixed-race pairs were generally associated with increased activity in the dACC and regions of dorsal lPFC, relative to judgments of same-race Black face pairs. In addition, an interesting pattern of activity appeared when comparing mixed-race judgments of athleticism with judgments of potential friendship. When judging whether a Black or a White individual was more athletic (vs. a potential friend), participants exhibited greater activity in dorsal and ventral regions of the lPFC, but little activity in the medial PFC, as in previous studies of prejudice control. By contrast, when judging between a Black vs. White face as a potential friend (vs. being athletic), strong activations were observed in middle region of the medial PFC that has been associated with processing of more familiar others and self-relevant stimuli. Although this area of mPFC is often interpreted in terms of social information processing, recent work by Amodio et al. (2006; see also Amodio & Frith, 2006) suggests that activity in this region is important for regulating one's social behavior according to the expectations of social norms. The lPFC activations that were observed when judging athleticism of mixed-race face pairs are consistent with the idea that response regulation did not involve personal interest (as in the friendship judgments), but rather the control of objective, impersonal responses. Additional research will be needed to further unpack that possibility that medial and lateral regions of the PFC are involved in different aspects of self-regulation when making social judgments.

It is notable that recent advances in understanding the role of the lPFC in the regulation of prejudice have been made using EEG, and these findings may aid in interpreting the fMRI results reviewed above. A large body of literature has suggested that left vs. right asymmetries in lPFC activity are associated with approach vs. withdrawal motivation. Amodio, Devine, & Harmon-Jones (in press) used EEG to measure changes in the lPFC after participants realized they had responded in a prejudiced manner and while they were given an opportunity to engage in an activity designed to reduce their level of prejudice. Indeed, the authors observed a reduction in left lPFC when participants believed they had acted in a prejudice way, and this reduction in activity was associated with high levels of guilt. However, when given a chance to make up for their transgression by reading magazine articles on how to reduce prejudice, lPFC activity was increased. Importantly, participants' self-reported desire to engage in prejudice-reduction activities predicted their degree of lPFC activity, whereas their desire to engage in other activities that were not related to prejudice reduction was not related to PFC activity. Although this research did not use fMRI, it is the first study to provide providing direct evidence that changes in the lPFC are associated with self-regulation in the context of racial prejudice.

Inhibition of race-biased emotion

Most neuroscience research on control has focused on mechanisms involved in the regulation of behavior. More recently, researchers have begun to investigate mechanisms for regulating one's affective responses to race. Lieberman et al. (2005) used fMRI to examine the neural processes underlying the control of race-related affect. In this study, participants completed a matching task while their brains were scanned. In one condition, participants saw a target face at the top of the screen and two additional faces at the bottom of the screen (Figure 5, Panel A). Their task was to choose which of two faces most closely matched the target face. This

condition was referred to as "perceptual encoding," because it involved matching one's visual image of the two faces. Target faces consisted of either White and Black male faces or colored shapes. In the case of faces, matches were determined no the basis of race. In a second "verbal encoding" condition, participants were presented with a target face at the top of the screen and the labels "Caucasian" and "African American" in the bottom of the screen (Figure 5, Panel B). Participants chose the label that best matched the target stimulus. Lieberman et al. (2005) reasoned that the process of encoding a face into a verbal representation involved the down-regulation of any emotional responses that might have been activated by the target stimulus (see Hariri, Bookheimer, & Mazziotta, 2000). As predicted, perceptual encoding of the targets produced greater amygdala activity to Black than White target faces, whereas this effect was absent in the verbal encoding condition. Instead, verbal encoding of the Black targets was associated with activity in ventrolateral PFC. The magnitude of this PFC response was inversely associated with amygdala activity, supporting the idea that ventrolateral PFC activity may play a role in regulating amygdala responses to Black targets and negative affect more generally (Lieberman, in press).

Neural basis of intergroup person perception

Most neuroscience studies of race bias from a social psychological approach have focused primarily on elucidating the automatic and controlled components on stereotyping and prejudice. However, researchers coming from a cognitive neuroscience perspective have emphasized the more basic role of person perception – how do we determine whether someone is part of our group? Neuroimaging research in this area suggests that medial regions of the PFC play an important role in several aspects of person perception and in the processing of social information (Amodio & Frith, 2006; Mitchell, Macrae, & Banaji, 2006) .

Neural substrates of ingroup vs. outgroup perception

The most basic form of social cognition involves determining whether an object as agentic (e.g., human) and distinct from the self. A large body of research has examined the neural correlates of *mentalizing*: the process of ascribing a unique perspective to another individual (Frith & Frith, 1999; Premack & Woodruff, 1978; Saxe, Carey, & Kanwisher, 2004). Theory of Mind (ToM) refers to the ability to mentalize, and several different tasks have been used to study mentalizing and ToM processes. In ToM cartoon studies, participants view and/or read cartoons that require one to take a character's unique perspective into account. Compared with cartoons that do not require perspective-taking, ToM cartoons typically elicit activity in a dorsal region of the mPFC located in Brodmann's Area (BA) 9/32 (Fletcher et al., 1995; Gallagher et al., 2000). Across several studies using a range of tasks, the act of mentalizing has been associated with activity in the same general region of dorsal mPFC (Saxe et al., 2004).

Mitchell and his colleagues have conducted several studies examining the neural substrates of social vs. non-social aspects of person perception (Mitchell, Heatherton, & Macrae, 2002). Commonly-used tasks in this line of research requires subjects to make judgments about an unfamiliar person that involves either social or non-social inferences. For example, in a study by Mitchell et al. (2002), participants viewed a series of noun-adjective pairs. Nouns were either names of people or inanimate objects, and adjectives could either describe a person (but not the object) or the object (but not the person). Mitchell et al. (2002) were interested in how patterns of brain activity differed on trials associated with a person-related judgment compared with judgments of inanimate objects. Across studies, social inferences were associated with increased activation in dorsal mPFC compared with non-social judgments (Mitchell, Banaji, & Macrae, 2005; Mitchell, Macrae, & Banaji, 2005, 2006). The region of activity associated with social

perception is similar to the region linked to mentalizing. Thus, the dorsal mPFC appears to be involved in perceiving a person as a social being. Some have argued that this process may form the basis of prejudice (e.g., Qui, 2006)

Research examining neural correlates of self-reflection suggest that thinking about one's own personality traits, compared with traits of a familiar but unrelated person (e.g., the president) is linked to activity in the middle mPFC (BA 10/32; Kelley et al., 2002). Subsequent work has shown that this region of mPFC is more active when thinking about either the self or a similar other compared a dissimilar other (Gobbini, Leibenluft, Santiago, & Haxby, 2004; Mitchell et al., 2006; but see Heatherton et al., 2006). By comparison, thinking about a dissimilar other is associated with activity in the dorsal mPFC. Thus, these findings suggest potential differentiation in the neural correlates of ingroup vs. outgroup perception. To date, fMRI research has not examined this effect within the context of racial prejudice, although there is reason to believe that similar effects would be observed.

Investigations of the social perception of similar vs. dissimilar others indicates that there may be important differences in the way we process information about members of the ingroup vs. the outgroup. However, additional research is needed to understand the meaning and implications of these different neural patterns. To date, research on person perception and social cognition have been rather descriptive, in that they have documented distinct patterns of brain activity for social and non-social judgments. As this line of work expands, researchers will begin to focus more on the functional properties of activations associated with social processes, such as their implications for the regulation of social behavior. Finally, it will be important for researchers to more fully integrate the findings from fMRI experiments with the rich body of theoretical and empirical work on intergroup processes in and social psychology literature. In all,

fMRI research on person perception, mentalizing, and the mirror neuron system stands poised to make important contributions to our understanding of prejudice and intergroup relations.

Neural basis of outgroup empathy

Most fMRI studies of social cognition have focused on the most basic process of perceiving a person as sentient entity with his or her own unique mental contents. Harris and Fiske (2006) have extended this line of inquiry to address how neural activity in these person-perception areas relate to specific qualities ascribed to members of different social groups, as suggested by the Stereotype Content Model (SCM; Fiske, Cuddy, Glick, & Xu, 2002). The SCM, proposes that the perception of social groups is primarily driven by evaluations along two independent dimensions: warmth and competence. Fiske et al. (2002) argued that the people's emotional reactions to different groups are associated with these factors. For example, groups defined by high warmth and high competence, such as middle class Americans and Olympic athletes, are associated with pride. Groups defined by high levels of warmth but low competence, such as the elderly and disabled, are described as pitiful. Highly competent/low-warmth groups, such as the wealthy, are met with envy. And most importantly for the present set of concerns, groups associated with low warmth and low competence – the homeless, the poor, African Americans and Hispanics – are met with disgust (Fiske et al., 2002).

Harris and Fiske (2006) used fMRI to determine whether judgments of warmth and competence were related to neural activations in regions linked to mentalizing and person perception. During scans, participants viewed pictures of people belonging to groups from each of the four quadrants of the SCM model. The authors observed significant mPFC activations relative to baseline when participants viewed pictures of groups associated with pride, envy, and pity. These activations were primarily in the middle region of the mPFC, suggesting that these

groups were processed similarly as the self. By contrast, groups associated with disgust did not elicit activity in this region. Harris et al. (2006) interpreted that lack of activity in this area as indicating *dehumanization* of these groups (see also Haslam, 2006). That is, low warmth/low competence groups were not being perceived as agentic human beings, but were rather perceived as inhuman objects, at least in terms of social emotional processing in the brain.

The results of Harris and Fiske (2006) suggest that prejudice and discrimination toward members of stigmatized social groups, such as African Americans, may be in part driven by a lack of "humanization" in some observers' social perceptions. However, the extant research suggests that the role of the mPFC in racial prejudice is more complex. For example, Wheeler and Fiske (2005) found that the categorization of Black (vs. White) faces elicited activity in the amygdala as well as the insula, a region implicated in visceral states and disgust, yet no difference was observed in the mPFC (L. Harris, personal communication). Thus, it appears that an understanding of the neural mechanisms of prejudiced person perception will require a consideration of a broad range of processes, of which mentalizing is just one. It is also worth noting that previous research has not been designed specifically to examine the role of mentalizing and racial prejudice – that is, to elicit mentalizing toward racial ingroups vs. outgroups – and so the jury is still out on this issue.

What have we learned about prejudice from fMRI studies?

Advances in neuroimaging methods have provided social psychologists with powerful new tools for studying the mechanisms of prejudice and discrimination. But has fMRI led to any significant new theoretical discoveries? This is a legitimate question often asked by many social psychologists. fMRI research on social processes is valuable in two general ways. First, there is value in the basic endeavor of brain mapping in order to begin to understand the functions of

different neural structures. The brain is a complex organ with much uncharted territory, and the only way to learn how it works is by observing activity as participants perform different types of tasks. Although there are caveats with this approach – neural operations are complex and specific structures often serve multiple and distributed functions (Poldrack, 2006) – it nevertheless serves an important role in cognitive neuroscience. Ultimately, brain-mappers hope to build a catalog of task-related activations that, over time, show consistent and coherent patterns of mental function.

The second way in which fMRI research is valuable is in elucidating mechanisms involved in psychological processes that cannot be inferred from behavior or are difficult to distinguish using the traditional tools of social cognition. In addition, the use of fMRI permits researchers to connect the social psychology literature on humans with the vast neuroscience literature on animals, opening the door for the crosstalk between fields and the application and integration of theoretical models from the two broad disciplines. From the prejudice researchers' perspective, the application of animal neuroscience models to questions of race bias may provide important information about how particular mechanisms involved in prejudice, stereotyping, and discrimination maybe interconnected. It is through these applications that fMRI research has contributed the literature in prejudice and stereotyping research. Here, we give a few examples of such contributions.

Patterns of behavior that would become known as implicit prejudice were first observed in the early 1980's (e.g., Gaertner & McLaughlin, 1983; Dovidio et al., 1986; Devine, 1989), and by the year 2000, implicit prejudice was a highly-replicated and established phenomenon. Yet social psychology lacked a theoretical explanation for what it was. Was it a cognitive association? Was it an emotion? How could it actually be unconscious? How might it influence behavior? How was it learned? How could it be unlearned? Although much research was aimed

at addressing these questions, there wasn't a theoretical foundation for how to conceptualize the process of implicit prejudice. The fMRI research linking implicit prejudice effects to the amygdala was groundbreaking in that it provided a concrete theoretical basis for the phenomenon. Through this work, we have learned that implicit prejudice likely involves a passive-learning memory system sensitive to affective cues (e.g., threats or punishments). It does not likely reflect conceptual representational networks as suggested by many social cognitive accounts. The social neuroscience research has shown that implicit prejudice is part of a subcortical response network that processes information rapidly and interfaces strongly with autonomic and behavioral systems. Moreover, linking implicit prejudice to the amygdala has allowed researchers to take the volumes of information gained from animal research on amygdala-based learning and memory and apply it to our understanding of how implicit prejudice may be learned and unlearned. For example, the unlearning of a classically-conditioned response involves a very different process than have been suggested by social cognition models that assume an associative learning process (e.g., Smith & DeCoster, 2000; see Amodio & Devine, 2006). These developments represent a major leap forward in our theoretical understanding of implicit prejudice.

A broader issue that is raised by the neuroscience approach to studying implicit prejudice concerns the meaning of the term. When implicit prejudice was first linked to amygdala activity, both prejudice and the amygdala were believed to reflect a fear response (Phelps et al., 2000). Over time, implicit prejudice still appears to be associated with amygdala activity. Yet researchers' interpretations of amygdala activity have changed. Currently, most researchers interpret patterns of amygdala activity as being associated with arousal or the emotional intensity of a stimulus, but not valence or fear per se (Anderson et al., 2002; Cunningham et al., 2004). To

the extent that the amygdala is the primary neural substrate of implicit prejudice, these more recent findings suggest that implicit prejudice may be better conceived as reflecting the intensity of one's reaction to an outgroup (vs. ingroup) face. Furthermore, neuroscience analyses of the amygdala and implicit prejudice force researchers to take a closer look at what participants are thinking while viewing faces of Black and White individuals. Although social psychologists go to great lengths to hide the true nature of the study from participants, anyone who has completed more than a few trials on an implicit prejudice task, such as the IAT, knows that the study is examining prejudice toward Black people. Participants completing an implicit prejudice task may become more vigilant for the presentation of Black faces or have stronger reactions when Black faces appear simply because they know that their responses to Black (vs. White) faces are being monitored. Thus, it is unclear whether the amygdala activity is related to participants' prejudiced reaction to the face, as typically inferred, or their anxiety about being in a prejudice study (although studies showing amygdala effects to subliminal pictures may argue against this alternative explanation). The role of anxiety in measures of implicit prejudice is an important one that will need to be resolved in future research.

As a second example, researchers have long distinguished between prejudice and stereotyping. But until recently, there was not a theoretical framework to specify the nature of their differences. It was unclear whether prejudice and stereotyping differed at the implicit level, and further unclear how either process might interface with behavior. A major obstacle to distinguishing between implicit prejudice and stereotyping is that they tend to operate in concert. That is, it is very difficult to design behavioral tasks capable of measuring these processes independently because they tend to be activated simultaneously. On the basis of neuroscience research regarding different regions of the brain involved in implicit affective vs. semantic

processing, we used fMRI to assess the activation of implicit prejudice and implicit stereotyping independently as they co-occurred (Potanina et al., 2006). By applying what is known about the different profiles of these neural regions, including their patterns of connectivity throughout the brain, we can develop a more concrete theoretical framework for how each process is learned, unlearned, expressed in behavior, and controlled. For example, our findings suggest that different prejudice reduction techniques are needed to target these two types of implicit bias, and that it may be best to use both types of reduction techniques in conjunction in order to most effectively diminish bias. Importantly, these advances were only possible through the integration of the social psychological and neuroscience literatures and the use of fMRI.

Finally, it is important to note that behavioral researchers can benefit from the advances made by fMRI research without using fMRI themselves. That is, new theoretical hypotheses about intergroup processes suggested by neuroimaging research can often be tested using behavioral methods (e.g., Amodio & Devine, 2006). Indeed, the range of behavioral tasks that may be used in the fMRI environment are limited, primarily due to the logistics of being confined to as small space and to the need to keep one's head very still. Therefore, behavioral studies are often the preferred way to test hypotheses suggested by fMRI research, particularly when they pertain to social behavior that is best studied in real-life interpersonal interactions. A major goal of this chapter is to convince researchers who are not interested in doing their own fMRI studies that there is value in considering the neuroscience literature in order to enrich behavioral approaches to the study of social behavior.

Conclusion

As research on prejudice and intergroup relations continues to evolve, researchers are increasingly integrating theories and methods of traditionally disparate fields such as cognitive

neuroscience. fMRI research is among the most recent approaches to be incorporated into the purview of prejudice research. Although relatively new, the fMRI approach to prejudice research in flourishing, and it has already begun to yield significant advances within social psychological theorizing. However, the advances described in this chapter are just the tip of the iceberg. Before long, the findings of fMRI studies on prejudice will be considered to be part of the cannon, and fMRI will move from having the status of a new trend to being another valuable tool in the prejudice researchers' box. By that time, chapters devoted to fMRI studies of prejudice will be a thing of the past, and the findings from neuroimaging research will be defined by their conceptual contributions. Until then, researchers who use fMRI to study prejudice should continue to look to unresolved theoretical issues that might be advanced by fMRI in ways that behavioral methods have not produced conclusive results.

Figures Captions

1. The MRI scanner. MRI scanning requires that the participant's head is centered inside the bore of a large electromagnet. Participants lie in a supine position on the scanner bed, and the bed is then moved into position. In addition, small movements may create problematic artifacts in the MR images. These restrictions of positioning and the need to remain extremely still during scans limits the types of tasks that can be used in experiments and may also affect the psychological experience of the participant. These limitations present special challenges for researchers interested in social behavior, such as prejudice researchers.

2. The amygdala comprises several small nuclei and is located bilaterally in the medial temporal lobe. The arrow on the left side indicates the anatomical image of the left amygdala. The arrow on the right side indicates functional activity of the right amygdala.

3. Lateral view indicating temporal lobe and PFC. Regions of dorsal and ventral lPFC have been associated with the controlled processing, and left PFC has been linked to semantic processes that play a role in stereotyping.

4. Medial view of the brain illustrating the dorsal ACC, dorsal mPFC, and middle mPFC. The shaded areas of these regions are those typically activated in studies of prejudice control and person perception described in the text.

5. Stimuli used by Lieberman et al. (2005) in their matching task. Panel A shows a sample stimulus of the perceptual encoding task, in which participants match a face with two comparison faces. Panel B shows a sample stimulus of the verbal encoding task, in which participants match a face with a verbal label.

References

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Fear and the human amygdala. *Journal of Neuroscience, 15*, 5879–5891.

Allport, G. W. (1954). The nature of prejudice.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652-661.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (in press). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (under review). Individual differences in responding without prejudice: The role of conflict detection and neural signals for control.

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

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology, 84*, 738-753.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*, 88-93.

Amodio, D. M., Kubota, J. T., Harmon-Jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs. external cues. *Social Cognitive and Affective Neuroscience, 1*, 26-36.

Amodio, D. M., & Potanina, P. V. (2006). Roles of the medial and lateral prefrontal cortex in regulating intergroup judgments. Manuscript in preparation.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., Gabrieli, J. D., & Sobel, N. Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6*, 196-202.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*, 170-177.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230-244.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype Activation and Control of Race Bias: Cognitive Control of Inhibition and Its Impairment by Alcohol. *Journal of Personality and Social Psychology, 90*, 272-287.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*, 1115-1118.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. Moskowitz (Ed.), *Cognitive social psychology: On the tenure and future of social cognition* (pp. 359-374). Mahwah, NJ: Erlbaum.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Breiter, H. C., Rauch, S. L., Kwong, K. K., Baker, J. R., & et al. (1996). Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. *Archives of General Psychiatry, 53*, 595-606.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*, 839-855.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83*, 1314-1329.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable Neural Components in the Processing of Black and White Faces. *Psychological Science, 15*, 806-813.

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and Explicit Evaluation: fMRI Correlates of Valence, Emotional Intensity, and Control in the Processing of Attitudes. *Journal of Cognitive Neuroscience, 16*, 1717-1729.

Davis, M., Hitchcock, J. M., & Rosen, J. B. (Eds.). (1987). *Anxiety and the amygdala: Pharmacological and anatomical analysis of the fear-potentiated startle paradigm*. San Diego, CA: Academic Press.

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*, 13-34.

Devine, P. G. (1989). Prejudice and stereotypes: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Devine, P. G., & Elliot, A. J. (1995). Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited. *Personality and Social Psychology Bulletin, 21*, 1139-1150.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22-37.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

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.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Fendt, M., & Fanselow, M. S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience & Biobehavioral Reviews, 23*, 743-760.

Fiske, S. T. (Ed.). (1998). *Stereotyping, prejudice, and discrimination*. New York: McGraw-Hill.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*, 109-128.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science, 286*, 1692-1695.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87-115.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46*, 23-30.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of the mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*, 11-21.

Gilbert, D. T., Fiske, S. T., & Lindzey, G. (Eds.). (1998). *The handbook of social psychology, Vol. 1* (4th ed.). New York, NY: McGraw-Hill.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., et al. (2006). Functional Specialization within Rostral Prefrontal Cortex (Area 10): A Meta-analysis. *Journal of Cognitive Neuroscience, 18*, 932-948.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, 1628–1635.

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

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

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*, 1–20.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *Neuroreport: For Rapid Communication of Neuroscience Research, 11*, 43-48.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups. *Psychological Science, 17*, 847-853.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H. k., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport: For Rapid Communication of Neuroscience Research, 11*, 2351-2355.

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review, 10*, 252-264.

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience, 1*, 18–25.

Henderson-King, E. I., & Nisbett, R. E. (1996). Anti-Black prejudice as a function of exposure to the negative behavior of a single Black person. *Journal of Personality and Social Psychology, 71*, 654-664.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*. Sunderland, MA: Sinauer Associates Inc.

James, W. (1890). *The principles of psychology*. Ny: Henry Holt and Company.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self?: An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*, 785-794.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377-395.

LeDoux, J. E. (1992). Emotion and the Amygdala. In J. P. Aggleton (Ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (pp. 339–351). New York: Wiley-Liss.

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58*, 259-289.

Lieberman, M. D. (in press). Why symbolic processing of affect can disrupt negative affect: Social cognitive and affective neuroscience investigations. To appear in A. Todorov, S. T. Fiske, & D. Prentice (eds.) *Social Neuroscience: Toward understanding the underpinnings of the social mind.* Oxford University Press.

Lieberman, M. D., Eisengerger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (in press). Putting Feelings into Words: Affect labeling disrupts amygdala activity to affective stimuli. *Psychological Science*.

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.

Lippman, W. (1922). *Public opinion*. New York: Macmillan.

Mackie, D. M., & Smith, E. R. (1998). Intergroup relations: Insights from a theoretically integrative approach. *Psychological Review, 105*, 499-529.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435-442.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience, 17*, 1306-1315.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, 99*, 15238-15243.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex, *NeuroImage, 26*, 251-257.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*, 655-663.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., & et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*, 1215-1229.

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

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*, 59-63.

Potanina, P. V., Pfeifer, J. H., Lieberman, M. D., & Amodio, D. M. (2006). *Distinct neural substrates of implicit prejudice and stereotyping*. Manuscript in preparation.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Qui, J. (2006). Neuroimaging: Peering into the root of prejudice. *Nature Reviews Neuroscience 7*, 508-509.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience, 6*, 1323-1328.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience, 15*, 1160-1175.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*, 79-82.

Rydell, B. J., & McConnell, A. R. (in press). Understanding Implicit and Explicit Attitude Change: A Systems of Reasoning Analysis. *Journal of Personality and Social Psychology*.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87-124.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190.

Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 297–315). San Diego, CA: Academic Press.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108-131.

Squire, L. R., & Zola, S. M. (1996). Ischemic brain damage and memory impairment: A commentary. *Hippocampus, 6*, 546-552.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science, 7*, 177-188.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*, 411-418.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling Racial Prejudice: Social-Cognitive Goals Affect Amygdala and Stereotype Activation. *Psychological Science, 16*, 56-63.

Wible, C. G., Han, S. D., Spencer, M. H., Kubicki, M., Niznikiewicz, M. H., Jolesz, F. A., et al. (2006). Connectivity among semantic associates: An fMRI study of semantic priming. *Brain and Language, 97*, 294-305.

Wig, G. S., Grafton, S. T., Demos, K. E., & Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nature Neuroscience, 8*, 1228-1233.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.

Table 1. Independent processes involved in intergroup bias and their associated neurocognitive function and neural correlates.

Role in intergroup bias	Neurocognitive function	Candidate structure(s)	Selected references
Implicit evaluative bias	Classical fear conditioning;	Amygdala	Phelps et al. (2000); Amodio et al. (2003); Cunningham et al. (2004)
	arousal; vigilance		
Implicit stereotyping	Conceptual priming	Temporal cortex & left lPFC	Potanina et al. (2006)
Detecting bias and need for control	Conflict monitoring	Anterior cingulate cortex	Amodio et al. (2004); Cunningham et al. (2004)
Inhibition of implicit prejudice	Response inhibition; Affect inhibition	Ventral lPFC	Lieberman et al. (2005); Cunningham et al. (2004)
Implementation of intended response	Regulative control	Dorsal lPFC	Cunningham et al. (2004); Richeson et al. (2003)
Outgroup perception	Mentalizing; Theory of Mind	Dorsal mPFC (BA 9/32)	Mitchell et al. (2005; 2006); Amodio & Frith (2006)
Ingroup perception	Processing of self and similar others	mPFC (BA 10/32)	Gobbini et al. (2004); Harris & Fiske (2006); Mitchell et al. (2006)
Detecting external cues for engaging control	Regulating behavior to external social cues	mPFC, rostral paracingulate	Amodio et al. (2006)

Note. lPFC = lateral prefrontal cortex; mPFC = medial prefrontal cortex.

Figure 1

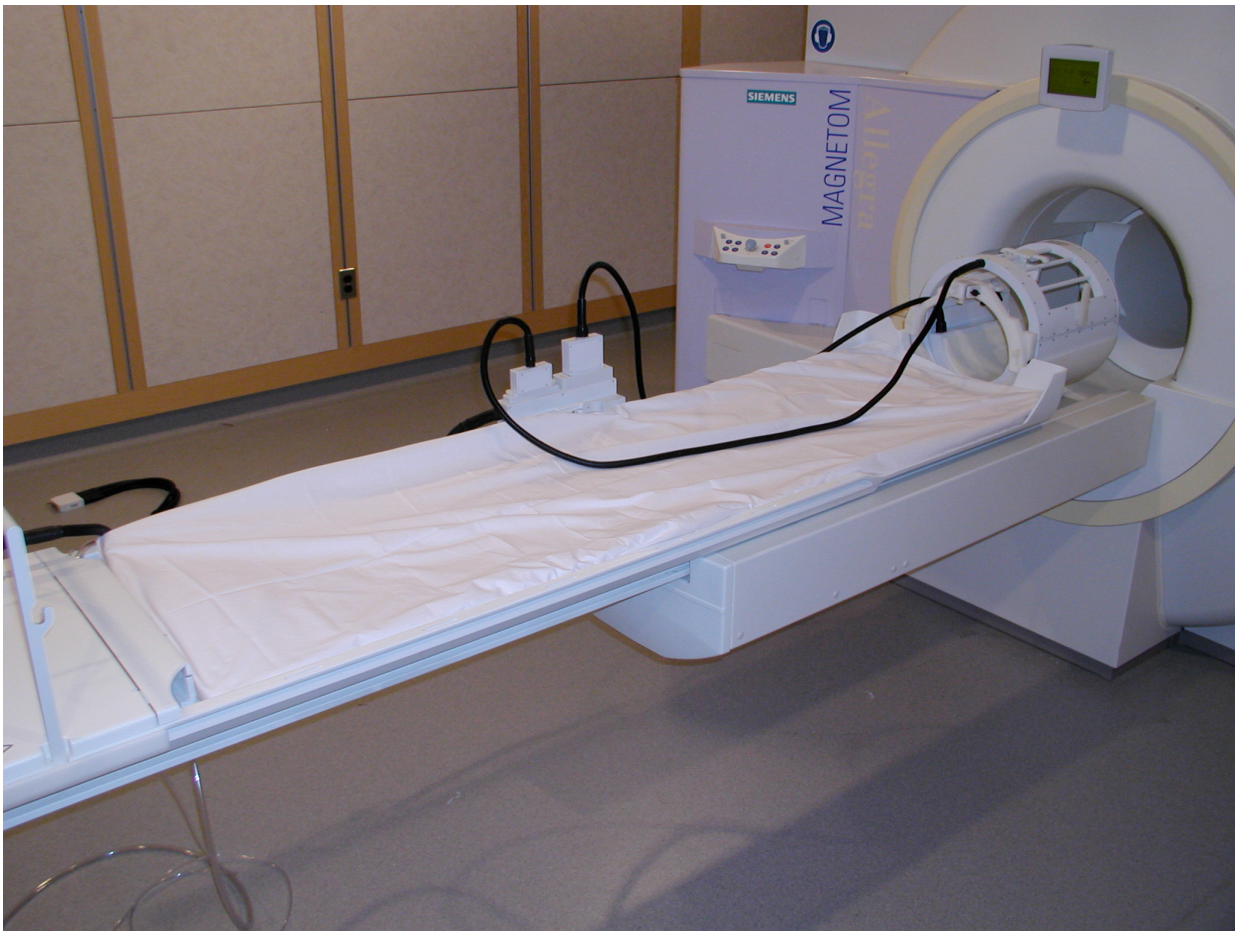


Figure 2

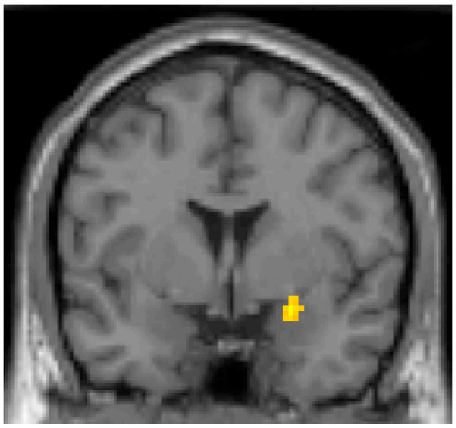


Figure 3

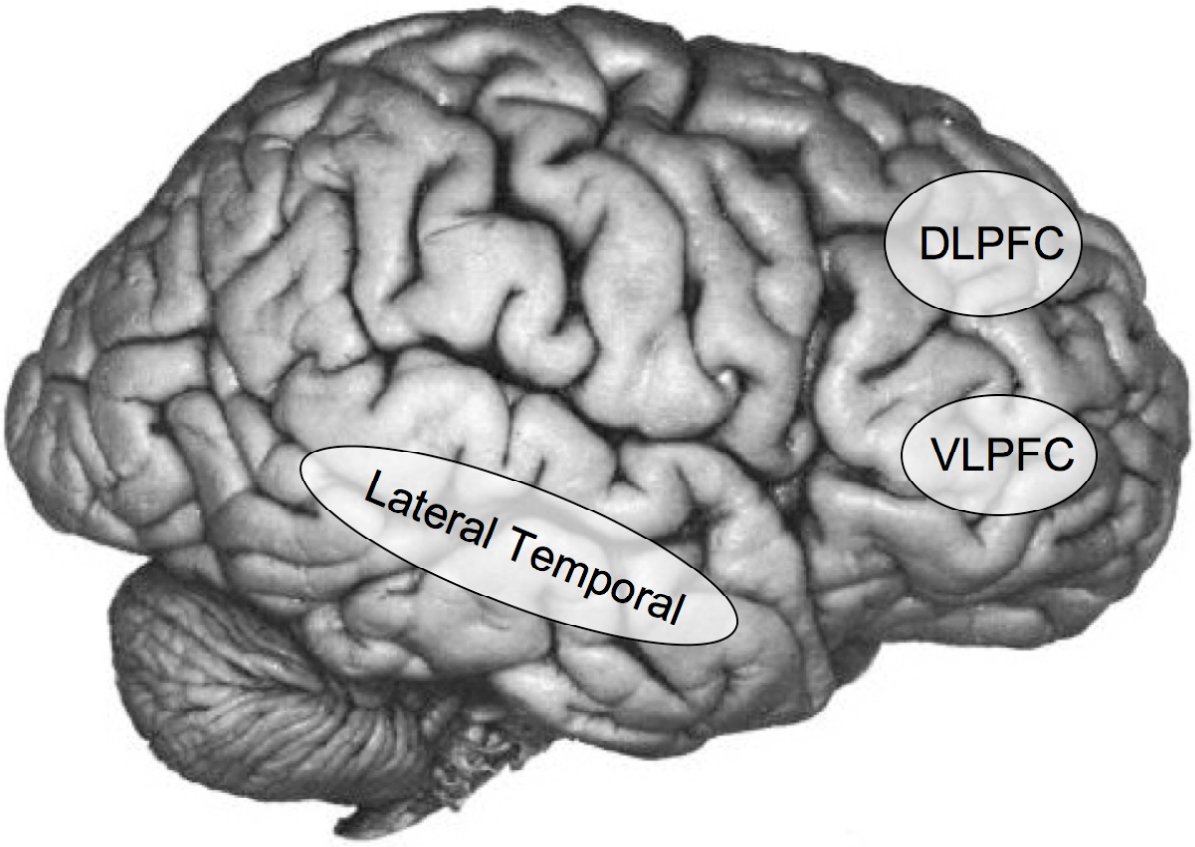


Figure 4

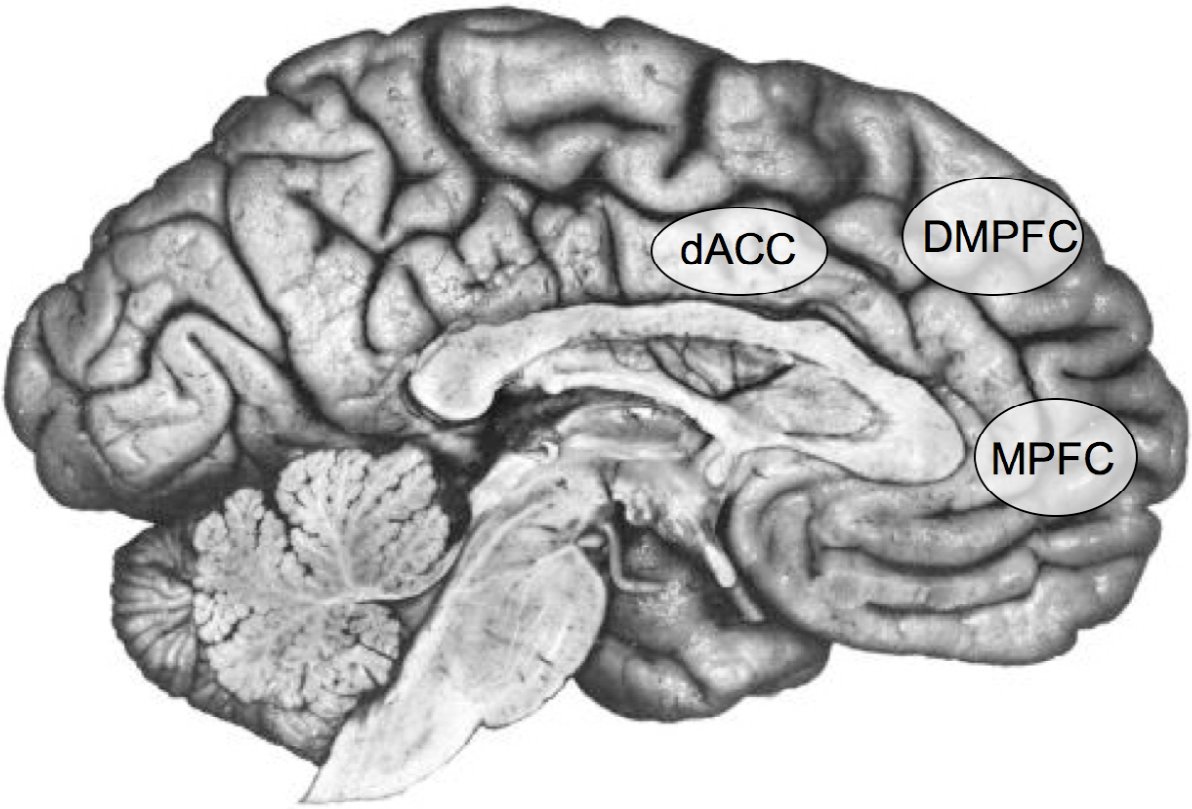
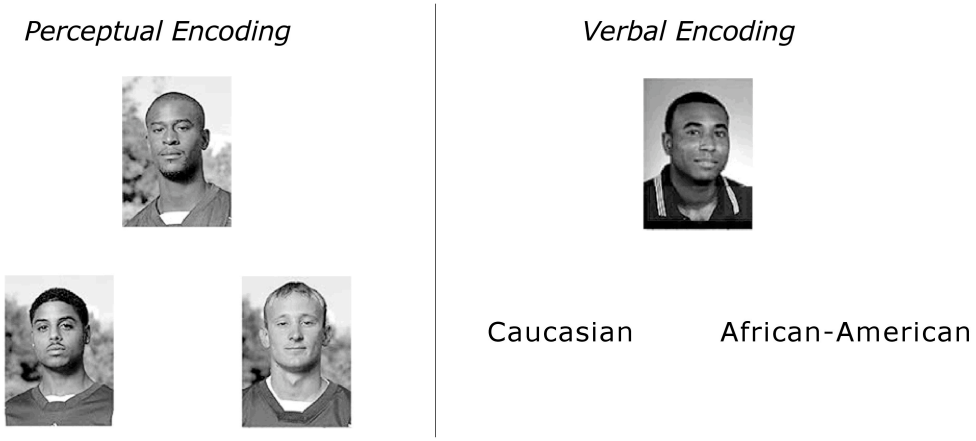


Figure 5


Perceptual Encoding
Verbal Encoding
Caucasian
African-American