EUROPEAN REVIEW OF SOCIAL PSYCHOLOGY
2008, 19, 1 – 54

Psychology Press
Taylor & Francis Group

The social neuroscience of intergroup relations

David M. Amodio
New York University, NY, USA

The social neuroscience approach integrates theories and methods of social psychology and neuroscience to address questions about social behaviour at multiple levels of analysis. This approach has been especially popular in the domain of intergroup relations, in part because this area of research provides a rich context for connecting basic neurocognitive mechanisms to higher-level interpersonal, group, and societal processes. Here I provide a brief description of the social neuroscience approach, and then review research that has used this approach to advance theories of (a) implicit racial bias and their effects on behaviour, (b) the self-regulation of intergroup responses, and (c) prejudice reduction. I also describe how the social neuroscience perspective suggests some important refinements to theoretical conceptions of implicit bias, prejudice control, and prejudice reduction.

Keywords: Social neuroscience; Intergroup; Stereotyping; Prejudice; Discrimination; Implicit bias; Self-regulation; Control; Prejudice reduction; Dehumanisation; Amygdala; Medial frontal cortex

In his preface to the first edition of *The Nature of Prejudice*, Allport weighed the complexities of human prejudices against the technological advances of the time. He mused that although it took ‘‘years of labor and billions of dollars to gain the secret of the atom, it will take still a greater investment to gain the secrets of man’s irrational nature’’ (1954, p. xvii). Although questions certainly persist with regard to human nature, Allport might be pleased to learn that those billions of dollars spent on developing atomic physics are now benefiting the study of intergroup bias in the new area of social neuroscience.

Social neuroscience is an approach to psychological research that integrates models of neuroscience and social psychology to study the

Correspondence should be addressed to David M. Amodio, Department of Psychology, New York University, New York, NY 10003, USA. E-mail: david.amodio@nyu.edu

http://www.psypress.com/ersp DOI: 10.1080/10463280801927937

mechanisms of social behaviour. As with any new approach, the emergence of social neuroscience has elicited both excitement and scepticism—excitement about connecting social processes to biological mechanisms and acquiring an expanded palette of methodologies, and scepticism that social neuroscience is a fad, gaining undue attention for its use of technology without producing important theoretical advances. Indeed, both reactions are warranted to some extent, and so the challenge for consumers of social neuroscience research is to be able to identify the important theoretical breakthroughs emerging from this broad and rapidly developing area. The goal of this review is to introduce readers to the social neuroscience approach to intergroup bias and to highlight ways in which this approach has begun to shed new light on long-standing questions about prejudice, stereotyping, and discrimination.

The social neuroscience approach has been particularly popular within the field of intergroup relations. Although the approach is new, research conducted in this area to date has been applied primarily to traditional issues in the field, serving to address long-standing questions regarding the implicit nature of racial stereotypes and evaluations, their expression in behaviour, the mechanisms through which intergroup responses are regulated, and methods for the reduction of bias. In addition, the integration of neuroscientific models is beginning to suggest new ways of looking at processes of self-regulation that extend beyond the classic sociocognitive perspective. After providing a brief introduction to the social neuroscience approach, I will review some of the major advances in the study of intergroup processes that it has produced.

THE SOCIAL NEUROSCIENCE APPROACH

In broad terms, social neuroscience refers to the study of the brain in the context of social processes. Initial applications of this approach focused on questions of "Where in the brain is . . . [insert process here]": such as, Where in the brain is implicit race bias? Where does control occur? What is the neural locus of mentalising (i.e., inferring the beliefs and intentions of others)? Of course, the idea that psychological processes operate in the brain does not in itself constitute an advance. However, these initial brain-mapping studies were necessary for establishing the basic layout of the brain as it relates to mechanisms of social behaviour. We now have a good working model of functional neuroanatomy that identifies basic structures for different aspects of learning and memory, vision, attention, action monitoring, and cognitive control, to name a few (for a comprehensive review, see Gazzaniga, 2004). Discerning the functions of underlying neural structures is particularly useful for determining whether a psychological phenomenon is associated with one or multiple underlying processes. As the field has evolved, the types of questions asked by researchers from both a neuroscience background and a social psychological background have moved from being primarily descriptive to being more theoretical and process-oriented. For example, a neuroscientist may be interested in how a consideration of social goals, relationships, and contexts enhance the understanding of neural function. For the social psychologist, the focus is often on how a consideration of neural structure and function may illuminate the understanding of neurocognitive mechanisms associated with social cognition and social behaviour, and how neuroimaging techniques may be used to measure processes not amenable to self-report or behavioural assessment.

With the promise of integrating neuroscience models and methods into social psychology come several potential caveats that are worth discussing, primarily concerning how brain activations are interpreted (Cacioppo et al., 2003; Poldrack, 2006). Most importantly, one cannot assume a one-to-one mapping of a psychological process onto a particular neural structure, nor can one assume that a particular experimental task elicits a process-pure instantiation of the phenomenon of interest. Take, for example, research on the amygdala, the small almond-shaped set of nuclei located bilaterally in the medial temporal lobes (Figure 1). Early brain-lesion studies in animals

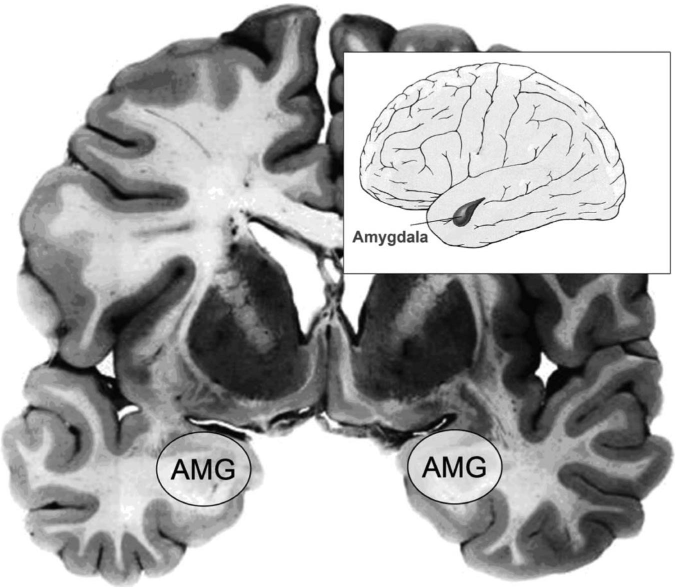

Figure 1. The amygdala (AMG) comprises a set of small nuclei and is located bilaterally in the medial temporal lobe, as shown in the coronal brain slice. The inset shows the position of the left amygdala as it would appear within the temporal lobe when viewed from the side.

discovered that damage to the amygdala impaired an animal's freezing behaviour in response to a classically conditioned stimulus—a hallmark of fear (LeDoux, 1992). On the basis of this work, the amygdala was believed to be the centre of fearful affect. In subsequent years, researchers armed with non-invasive neuroimaging techniques began to study amygdala function in humans, using a much wider array of experimental tasks. Sure enough, amygdala activity was increased during tasks that elicited fear (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998), and amygdala-lesioned patients showed selective impairments in fear processing (Adolphs, Tranel, Damasio, & Damasio, 1995).

However, interpretations of the amygdala began to get more complicated as studies found that it also became activated in response to positive stimuli and, more generally, when a participant was anticipating either a reward or punishment (Baxter & Murray, 2002; Paton, Belova, Morrison, & Salzman, 2006). Such findings required researchers to revise their functional interpretation of the amygdala from being a fear module to being a critical structure for vigilance, arousal, learning, and the orchestration of the fight-or-flight response—processes strongly engaged in fearful situations, but which do not represent fear per se (Davis & Whalen, 2001; Phelps, 2006; Whalen, 1998). As a result, the findings from a large body of amygdala research must be reinterpreted. In the realm of intergroup relations, research linking amygdala activation to implicit race bias initially concluded that implicit bias is rooted in subconscious fear processing (e.g., Amodio, Harmon-Jones, & Devine, 2003; Cunningham et al., 2004a; Hart et al., 2000; Phelps et al., 2000), but now that conclusion is less certain (e.g., Amodio & Devine, 2006). This issue continues to be problematic when researchers use brain activations as the primary outcome measure and interpret them as a specific psychological phenomenon (sometimes on the basis of past work using different tasks) without corroboration of behavioural measures or strong experimental designs. For example, amygdala activation is often interpreted as an outcome measure of "fear", despite the fact that much research suggests it may reflect somewhat different processes in different contexts. This problem is compounded when tasks used to elicit brain activity are not validated by corresponding self-report or behavioural evidence that could be used to corroborate a particular interpretation. This major interpretational problem is referred to as a "reverse inference" (Cacioppo et al., 2003; Poldrack, 2006). Given the complexity of processes typically studied by social psychologists, social neuroscience findings are particularly susceptible to the reverse inference fallacy.

At its best, social neuroscience research provides a bridge between theories of human social behaviour and the expansive literatures on human and non-human animal neuroscience. For example, by identifying the basic neural processes involved in prejudice and stereotyping, researchers may bring well-characterised animal models of learning, memory, and attention to bear on questions of how intergroup biases are learned, stored, expressed, regulated, and unlearned. In addition, the social neuroscience approach brings together methods used across a wide spectrum of disciplines, combining traditional social psychological research tools with assessments of brain activity, autonomic arousal, hormones, and immune responses, to inspire more comprehensive models of the mental and physical processes involved in social behaviour. In what follows, I describe how the social neuroscience approach has been applied to issues of implicit race bias, the self-regulation of prejudice, and prejudice reduction, with special emphasis on how each of these processes interface with behaviour.

THE SOCIAL NEUROSCIENCE OF INTERGROUP RELATIONS

The social neuroscience approach has been especially popular in research on intergroup relations. On the surface, these traditionally disparate areas appear to make strange bedfellows. However, the field of intergroup relations in social psychology has a tradition of integrating multiple levels of analysis. Although much research in this field focuses on high-level phenomena such as intergroup attitudes, motivations, social identity, intergroup behaviours, and societal-level judgement and decision making, research in each area typically involves elements of more basic-level processes involving implicit thoughts, emotions, and behaviours, and the mechanisms through which they are regulated. It is through this more basic level of analysis that the neuroscience literature may interface with questions about higher-level intergroup phenomena. In particular, research on the sociocognitive substrates of implicit race bias and the regulation of intergroup responses provides a direct bridge to research in cognitive and affective neuroscience on related issues of automaticity and control. Furthermore, many of the central components of intergroup bias (e.g., the construct of implicit bias) are exceedingly difficult to study using the traditional methods of social psychology, as they appear to be impervious to introspection, and thus to self-report, and are difficult to exact through behavioural measurement. Many physiological methods can circumvent these issues by measuring neural indicators of bias directly. For these reasons, the social neuroscience approach has been particularly attractive and fruitful among psychologists studying intergroup relations.

Implicit racial bias: Elucidating the construct and its link to intergroup behaviour

The notion that racial biases may operate automatically in the unconscious mind was a major breakthrough in intergroup research that galvanised the field in several ways (Devine, 1989; Dovidio, Evans, & Tyler, 1986; Gaertner & McLaughlin, 1983; Greenwald & Banaji, 1995). This discovery explained why individuals who consciously reject prejudice nevertheless show evidence of bias in their non-deliberative behaviours. Furthermore, the idea of dissociable implicit and explicit components of racial bias promised to provide a powerful theoretical model for explaining how attitudes and beliefs shape discriminatory behaviours (see also Wilson, Lindsay, & Schooler, 2000). Whereas previous work had shown that explicitly reported racial attitudes are often poor predictors of behaviour (e.g., Crosby, Bromely, & Saxe, 1980), the hope was that assessments of implicit bias might prove more successful. However, after nearly two decades of research on implicit (vs explicit) racial biases, an understanding of the link between racial bias and behaviour remains elusive (Blair, 2001; Fazio & Olson, 2003). With few exceptions among hundreds of studies (e.g., Amodio & Devine, 2006; Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; McConnell & Leibold, 2001), implicit measures have not proven to be strong or reliable predictors of intergroup behaviour (in part because few studies have examined effects on behaviour).

What explains the poor correspondence between measures of implicit race bias and behaviour? Arguing from a social neuroscience perspective, I have proposed that the predominant theoretical conceptualisations of implicit social cognition may be limited in their ability to account for the full range of processes involved in implicit biases or the way that these processes are expressed in behaviour (e.g., Amodio, in press; Amodio & Devine, in press; Amodio et al., 2003). Extant models of implicit social cognition (e.g., Devine, 1989; Fazio, Chen, McDonel, & Sherman, 1982; Gawronski & Bodenhausen, 2006; Greenwald & Banaji, 1995; Smith & DeCoster, 2000) are rooted in associative (and some connectionist) models of semantic information processing (McCelland & Rumelhart, 1985; Shiffrin & Schneider, 1977). These models assume that stereotypes and evaluations are connected to social targets (e.g., African Americans) through a network of semantic associations. This view of social cognition posits that both stereotyping and prejudice operate according to a particular set of parameters. Like any theoretical perspective, this view has influenced the way psychologists conceive of implicit bias and how it is activated and controlled, as well as how researchers design experiments to study implicit intergroup phenomena. However, although this view of implicit social cognition provides a parsimonious account for how information is stored and activated in memory, it corresponds to only a subset of the major learning and memory systems that have been identified in the behavioural and cognitive neuroscience literatures, as illustrated in Figure 2 (Squire & Knowlton, 2000; Squire & Zola, 1996).

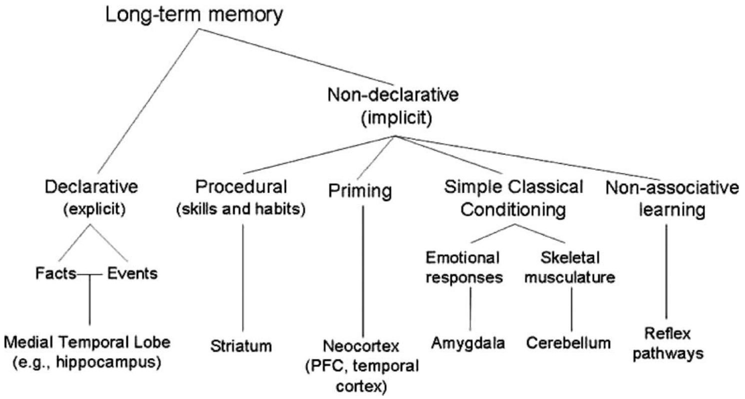

Figure 2. Diagram illustrating a framework of multiple memory systems and their respective neural substrates. Adapted from Squire and Zola (1996).

As shown in Figure 2, the broader literature on learning and memory suggests a major distinction between declarative (explicit) and non-declarative (implicit) forms of memory. Importantly, the literature indicates multiple distinct forms of implicit memory, each associated with different parameters of operation and with unique neural substrates. For the present set of issues, this model makes a critical distinction between memory systems for implicit semantic memory (i.e., conceptual priming) and affective systems of memory. Although social psychological theorising has generally assumed a single mode of implicit processing, the multiple-memory systems model suggests that it may be important to consider different forms of implicit processing. For example, intergroup theorists have long distinguished between stereotypes and prejudice, such that stereotypes refer to the content of semantic knowledge about a group, whereas prejudice refers to one's evaluation of the group (Allport, 1954; Dovidio, Brigham, Johnson, & Gaertner, 1996; Park & Judd, 2005). This descriptive distinction has been made for implicit as well as explicit levels of processing (Greenwald & Banaji, 1995; Wittenbrink, Judd, & Park, 1997). However, in research on implicit race bias the distinction between stereotyping and evaluation has not been clear, and researchers have generally assumed that implicit stereotyping and evaluation arise from the same underlying mechanism.

Although stereotypes and evaluations typically go hand in hand in outward expressions of behaviour, the multiple memory systems framework suggests that at their roots they arise from different underlying systems of semantic and affective memory, respectively. My colleagues and I have suggested that a consideration of the different functional properties of these two forms of memory may clarify our understanding of how these implicit biases operate in the mind and in behaviour (Amodio, in press; Amodio &

Devine, in press). In what follows, I describe these two different forms of implicit memory as they have been studied in the neuroscience and cognitive psychology literatures, and then discuss how this multiple-memory system framework may help to clarify some long-standing issues in research on implicit race bias.

Semantic associations. As noted above, predominant theories of implicit social cognition (including implicit racial bias) are rooted in associative models of information processing (e.g., Devine, 1989; Smith & DeCoster, 2000). These models correspond to the general rubric of semantic memory, and in the cognitive neuroscience literature this form of memory is often referred to as ‘‘conceptual priming’’ (Gabrieli, 1998). Like associative theories in cognitive psychology (e.g., McClelland & Rumelhart, 1985), models of conceptual priming assume that abstract concepts are represented in a parallel-distributed semantic network, and that the activation of a particular concept may activate or inhibit semantically related concepts in a spreading-of-activation manner (Gabrieli, 1998; Logan, 1990; Roediger & McDermot, 1993). These semantic associations are formed across repeated stimulus pairings in a probabilistic fashion, and thus learning in this system occurs relatively slowly, building over time (Poldrack, Selco, Field, & Cohen, 1999; Reber & Squire, 1994; Shiffrin, 2003). Similarly, the extinction of semantic associations occurs through repeated non-pairings; thus the process of learning and unlearning associations is assumed to be rather symmetrical.

Neuroimaging studies have consistently linked semantic priming effects with distributed regions of neocortex (Squire & Zola, 1996), including regions of left posterior dorsolateral prefrontal cortex (dlPFC, e.g., Blaxton et al., 1996; Demb et al., 1995; Raichle et al., 1994, Wagner, Gabrieli, & Verfaellie, 1997) and temporal cortex (Rissman, Eliassen, & Blumstein, 2003; Schacter & Buckner, 1998; Squire, 1992) (Figure 3). As with associative models of semantic networks, the neocortical substrates of conceptual priming appear to have limited connections to basic behavioural and autonomic systems. Rather, the distributed nature of semantic priming effects in neocortex suggests primary links to the higher-order processes of social cognition, self-reflection, and theory of mind (Amodio & Frith, 2006; Amodio, Kubota, Harmon-Jones, & Devine, 2006; Frith & Frith, 1999; Kelley et al., 2002; Mitchell, Banaji, & Macrae, 2005; Ochsner et al., 2005; Saxe & Kanwisher, 2003; Yonelinas, 2002). Neuroscience and cognitive psychology models of conceptual priming do not generally include prescriptions for how semantic associations influence behaviour. However, given the neural substrates for conceptual priming and their proximity to regions involved in higher cognition, one may infer that conceptual priming should influence judgements, perceptions, and semantic associations—processes that likely drive responses on self-report measures and on semantic priming tasks.

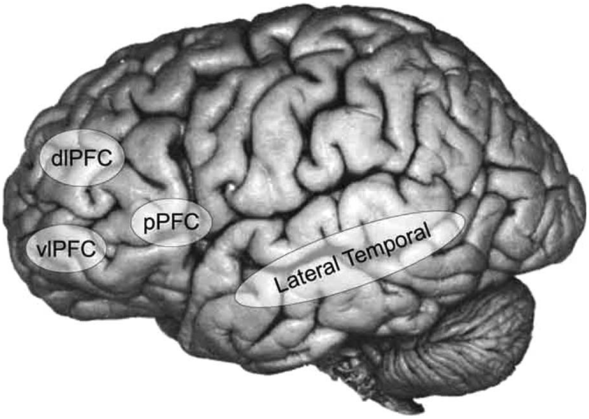

Figure 3. Lateral view of the brain. Labelled regions include the dorsolateral prefrontal cortex (dlPFC), ventrolateral prefrontal cortex (vlPFC), posterior prefontral cortex (pPFC), and temporal lobe. Note that conceptual priming has been associated with pPFC and temporal activations on the left side, whereas response inhibition has been associated with vlPFC activation on the right side. dlPFC activations associated with various forms of control have been observed on both sides.

Classical conditioning. Since Pavlov's (1927) famous observations, hundreds of studies have investigated the mechanisms of classical conditioning, a form of implicit affective memory (LeDoux, 2000). Many studies have examined *fear conditioning*—associating an aversive unconditioned stimulus (US) with a neutral conditioned stimulus (CS)—in animals such as rats (Davis, 1992) as well as humans (e.g., Adolphs et al., 1995; Phelps & LeDoux, 2005). A key feature of classical conditioning is that it is acquired rapidly, often after a single US–CS pairing (LeDoux, 1996). This style of learning stands in contrast to the learning of semantic associations, which occurs slowly over the course of many repeated pairings. Extinction of conditioned fear may occur after repeated exposure to the CS in a safe (neutral or appetitive) context. However, extinction occurs very slowly, if at all (Gale et al., 2004), and exposure to previously extinguished US–CS pairs results in rapid and stronger ''reconditioning'' (Bouton, 1994). Hence, classical conditioning is an extremely tenacious form of implicit memory that is less amenable to change than semantic associations. In this regard, the process of learning and unlearning implicit affective associations is asymmetrical.

As described above, neuroanatomical studies have established that classical fear conditioning is primarily dependent on the amygdala (Fendt & Fanselow, 1999; LeDoux, 1992). The amygdala is part of a set of ''rapid response'' structures activated and expressed within milliseconds of a potentially threatening event, such that sensory information is relayed by the thalamus via a single synapse to the amygdala for initial processing while slower, more elaborative processing continues throughout the cortex (LeDoux, 2000). For example, research has shown that the amygdala responds to threatening faces without participants' conscious awareness (Brieter et al., 1996; Morris et al., 1996). This "quick and dirty" detection quality makes fear conditioning an extraordinary mechanism for survival, but at the same time, relatively resistant to change and prone to generalisation. In animals with less-developed neocortices the amygdala is a primary mechanism for orchestrating adaptive behaviour, such as basic approach/ withdrawal responses and activation of the autonomic nervous system (e.g., to prepare the body for a fight or flight response). The amygdala and its associated subcortical structures accomplish this function through their strong connections to systems for initiating and monitoring behaviour, including brainstem structures, the thalamus, hypothalamus, basal ganglia, and medial prefrontal cortex (mPFC; Davis & Whalen, 2001). Because the amygdala is not known to process complex semantic information, it is less likely to influence higher-level judgements and decisions. Importantly, classical conditioning and semantic associations (i.e., retrieval or selection) are associated with different neural substrates, and research has demonstrated dissociations of classical fear conditioning and forms of semantic memory in brain-lesion patients (Bechara, Damasio, & Damasio, 1995; LaBar et al., 1995).

Although past research has not applied a multiple-memory systems approach to implicit race bias, several studies have linked implicit measures of evaluative racial associations to activity of the amygdala. For example, Phelps et al. (2000) used functional magnetic resonance imaging (fMRI)—a measure of blood flow in the brain—to examine White participants' amygdala activity in response to Black vs White faces. Researchers typically infer that increased blood flow to a specific brain region occurs because neurons in that region had fired and require replenishment by oxygenated blood. Thus, greater blood flow is used as an index of brain activity. In their study, Phelps et al. (2000) did not find a significant difference between amygdala activity to Black vs White faces. However, they found that that the relative difference in participants' amygdala response to Black vs White faces was associated with their scores on an Implicit Associations Test (IAT) of pleasant/unpleasant associations with Black vs White faces (i.e., implicit evaluative bias), and also with a startle-eyeblink index of amygdala response to Black vs White faces (see also Hart et al., 2000). However, these physiological measures of bias were not associated with participants' explicit racial attitudes. This work provided the first clues that implicit bias, as measured by reaction-time measures such as the IAT, may reflect activity of specific structures in the brain associated with affect.

Inspired by advances in affective neuroscience concerning the role of the amygdala (Davidson & Irwin, 1999; LeDoux, 1996), my colleagues and I became interested in whether implicit evaluative bias, as assessed by reaction-time measures such as the IAT, might reflect emotional responses, beyond the cognition-based evaluative associations that were traditionally assumed to drive these effects (Fazio et al., 1995; Greenwald, McGhee, & Schwartz, 1998). We designed research using the startle-eyeblink method to index amygdala activity in response to White, Black, and Asian faces as a way to measure emotional responses independently of ‘‘cognitive’’ associations (Amodio et al., 2003). The startle-eyeblink measure provides a temporally precise way to assess amygdala activity. The defensive eyeblink is one component of the whole-body reflex that occurs in response to a startling stimulus, such as a loud noise. Much research has shown that the magnitude of this startle response is modulated by one’s affective state at the moment one is startled (Lang, Bradley, & Cuthbert, 1990). For example, if a person is in a threatened state (e.g., watching a horror movie), then a loud noise might cause him to jump out his chair. But if he is in a pleasant or appetitive state (e.g., while watching a video of tasty foods), he would have a much milder reaction to the noise. Importantly, the affective modulation of the startle reflex is controlled by the amygdala and thus reflects amygdala activity (Davis, 1992; LeDoux, 1992).

Using electrodes to measure the contraction of the orbicularis oculi—the muscle surrounding the eye that contracts during a blink—one can assess changes in eyeblink magnitude to a startling noise as an indicator of amygdala activity at the moment the noise was administered. In our study, participants occasionally heard loud noise blasts (i.e., startle probes) through headphones while viewing White, Black, or Asian faces. By delivering startle probes either 400 or 4000 ms following the onset of a face picture, we could precisely measure the amygdala activity occurring very early vs later in face processing. In this regard, we were able to investigate the role of the amygdala in automatic vs potential controlled intergroup responses. By comparison, fMRI (at the time) required blocked experimental designs, such that participants would observe a block of several Black faces, followed by a block of several White faces, and an average estimate of amygdala activity for each block would be examined. Although Phelps et al. (2000) used the startle-eyeblink method as an additional index of amygdala activity, they did not assess startle responses occurring at very early latencies following face onset, and thus were unable to make strong inferences about the automaticity of the response. As such, little was known about the role of the amygdala (and affect more generally) in the automatic processing of outgroup faces.

A secondary goal in conducting this research was to test whether individual differences in people’s motivations to respond without prejudice were related to rapidly activated affective responses to Black faces (vs White or Asian faces). Our prior research had recently shown that participants’ implicit racial biases—which at that time were assumed to be inevitable and immutable—varied as a function of their motivations to respond without prejudice (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). That is, White Americans report that they respond without prejudice towards Black people for two independent reasons—in order to meet their personal, internal standards, and to avoid negative reactions from others who may disapprove of prejudice (i.e., normative, external standards; Plant & Devine, 1998). Thus, people may be motivated to respond without bias primarily for internal reasons, primarily for external reasons, for a combination of internal and external reasons, or they may not be motivated for either reason. Interestingly, Devine et al. found that people who were motivated only by internal reasons—that is, they personally rejected prejudice, regardless of what others thought—consistently exhibited lower levels of implicit racial bias on reaction-time measures than participants reporting each of the other motivational profiles. These results indicated that participants' motivational profiles could be used to distinguish between low-prejudice people who exhibited high vs low levels of implicit race bias. This was one of the first studies to show a consistent pattern of individual differences in scores on tests of implicit bias.

The results of Devine et al. (2002) raised questions about why only a subset of low-prejudiced White people (those responding primarily for internal reasons) exhibited low levels of implicit race bias on reaction time measures. Were they better at regulating their bias? Or did they genuinely have weaker negative affective associations with Black people? The startle-eyeblink method permitted us to test whether the same individual differences observed by Devine et al. (2002) would be found on a measure assessing (a) emotional reactions that (b) occurred within a few hundred milliseconds following the presentation of a Black vs White (or Asian) face. Our results showed a pattern of greater affective response to Black than White (or Asian) faces across participants, although this pattern was moderated by participants' internal and external motivations to response without prejudice. As in Devine et al. (2002), participants who were primarily internally motivated showed lower levels of bias at both short and long startle probe intervals, compared with participants motivated by a combination of internal and external concerns as well as those who were not internally motivated (i.e., those high in prejudice). These results provided evidence for a specifically affective form of automatic racial bias. Moreover, they suggested that individual differences in motivations to respond without prejudice moderated the degree of amygdala activity in response to faces, consistent with the findings of Devine et al. (2002).

Since these initial investigations of amygdala activity in response to ingroup vs outgroup faces, several other studies have also related amygdala activity to racial bias (Cunningham et al., 2004a; Lieberman et al., 2005; Ronquillo et al., 2007; Wheeler & Fiske, 2005; for a review, see Eberhardt, 2005). Together they suggest that implicit evaluative race bias is rooted in affective processes associated with the amygdala and, by extension, a classical fear conditioning mechanism.

Implications for implicit race bias. Research linking evaluative forms of race bias to the amygdala suggested that implicit race bias may reflect a combination of affective and semantic memory systems. Based on the operating characteristics of semantic association and classical fear conditioning, we can derive a set of predictions for how implicit evaluation and implicit stereotyping may differ in terms of learning, activation, regulation, expression, and extinction, as summarised in Table 1. Amodio and Devine (2006) provided an initial test of the predictions that individual difference measures of implicit stereotyping and evaluation should be independent and should predict different forms of race-biased behaviours. For example, the multiple memory systems model posits that implicit evaluation is driven largely by affective processes and, when activated, causes enhanced automatic nervous system activity and is expressed in non-verbals and withdrawal behaviours associated with threat. By contrast, implicit stereotyping is driven by semantic (cognitive) processes that, when activated, should be unrelated to automatic arousal, but should instead bias information processing and thus influence judgements and decision making in a stereotype-consistent way.

In order to test these basic predictions it was critical to use measures of implicit stereotyping and evaluation that do not confound evaluation with stereotype content. That is, most stereotypes of African Americans are negative in valence, and thus most expressions of implicit bias involve a combination of semantic and affective associations. To measure implicit evaluations in the absence of stereotypes, we used the typical version of the IAT (Greenwald et al., 1998) in which participants categorised White vs Black faces in combination with pleasant vs unpleasant words that were unrelated to racial stereotypes. This measure provides a straightforward assessment of evaluative associations irrespective of stereotype content.

TABLE 1
Predictions for how implicit evaluation and implicit stereotyping may differ

Parameter	*Implicit stereotyping*	*Implicit evaluation*
Acquisition	Probabilistic, slow	Single shot, fast
Activation	Automatic	More automatic? (just a few synapses)
Regulation	Behavioural override, lateral inhibition	Behavioural override
Extinction	Probabilistic, slow	Complicated, perhaps never
Expression	Person perception, social judgements	Autonomic and basic behavioural responses (e.g., non-verbals)

The greater challenge in our research was to measure implicit stereotyping in the absence of evaluations. To this end, we developed an IAT in which participants categorised faces as White vs Black and judged words associated with athleticism or (un)intelligence—two central dimensions of the African American stereotype (Devine & Elliot, 1995). In order for the intelligence and athletic words to be categorised along a single dimension (a technical requirement of the IAT), participants were instructed to classify words as "mental" vs "physical". In pretests, "physical" words (e.g., *basketball, dance, run*) were rated as more strongly associated with African than White Americans, whereas "mental" words (e.g., *educated, math, read*) were rated as more strongly associated with White than African Americans (Amodio & Devine, 2006). Both sets of words were rated as moderately positive and therefore could not be correctly classified on the basis of valence. In an initial study, participants completed both IATs in counterbalanced order. Participants exhibited significant effects of both implicit evaluation and stereotyping, replicating past work. More importantly, participants' scores on the two IATs were uncorrelated, thereby supporting our first hypotheses that implicit stereotyping and evaluation reflect independent underlying mechanisms.

Behavioural outcomes of implicit stereotyping vs evaluation. On the basis of the multiple-memory systems model, we hypothesised that implicit stereotyping and evaluation should predict different expressions of bias. We suggested it is possible that the low correspondence between implicit measures and behaviour in past work may have been a result of mismatched measures. Indeed, it is notable that the few studies in which a relationship between implicit bias and behaviour was observed have used measures of implicit evaluative bias, rather than implicit stereotyping, to predict interpersonal types of behaviours, in line with our hypothesis (e.g., Ashburn-Nardo, Knowles, & Monteith, 2003; Dovidio et al., 1997, 2002; Fazio et al., 1995; McConnell & Leibold, 2001; Wilson et al., 2000). However, past research has not directly compared the predictive power of different types of implicit bias.

In the first direct test of our predictions, participants read an essay ostensibly written by an African American college student (Moreno & Bodenhausen, 2001) and then, as per the cover story, inferred the writer's personality traits and their degree of friendliness. Participants then completed the evaluative and stereotyping IATs. The IATs were completed at the end of the session so that participants would not suspect the study was about racial issues while they made their earlier ratings of the African American writer. Again, participants' scores on the two IATs were not significantly correlated. However, a set of simultaneous regression analyses indicated that stereotyping IAT scores uniquely predicted more stereotype-consistent trait ratings of the African American writer, whereas evaluative

IAT scores uniquely predicted higher ratings of friendliness. Neither IAT predicted ratings of traits unrelated to racial stereotypes. These results added support to the notion that implicit stereotyping and evaluation represent independent underlying constructs. Furthermore, they provided the first evidence that different forms of implicit bias are expressed in different types of responses to outgroup members.

To provide a more ecologically valid test of our hypothesis, a third study examined the effects of implicit stereotyping vs evaluation on behavioural responses towards an African American person. This study took place in two separate phases. In an initial experimental session, participants completed the stereotyping and evaluative IATs. Several weeks later, the participants returned for what they believed was a separate study in which they would work with a partner on a set of tests, with the goal of achieving the highest combined score (to gain entry in a lottery for $50). Participants learned that the tasks tested their academic (math, verbal skills) and non-academic (sports, pop-culture knowledge) abilities, and that they would have to complete two of these tests and their partner would have to complete the other two. They also learned that their partner was an African American male. Before meeting with their partner, participants reported separate performance expectations for themselves and their partners. Next, participants waited outside the experimental room to meet the partner and to complete their tasks. While waiting, they were asked to sit in a row of eight chairs, the last of which held the African American partner's coat and backpack (Macrae, Bodenhausen, Milne, & Jetten, 1994). Although participants' scores on the two IATs were not correlated, a set of simultaneous regressions revealed that implicit stereotyping uniquely predicted lower performance expectations for the partner on academic (vs non-academic) tasks. By contrast, implicit evaluative bias uniquely predicted greater seating distance from the partners' belongings. Taken together, the three studies reported by Amodio and Devine (2006) provided evidence that implicit stereotyping and evaluation reflect independent underlying memory systems that arise from distinct neurocognitive processes and are expressed in different forms of behaviour. Moreover, these results suggest that greater correspondences between implicit measures and behavioural outcomes may be obtained when one considers that different forms of implicit bias may reflect distinct underlying neurocognitive systems that are expressed through different behavioural channels.[1]

[1] A potential limitation of the behavioural studies by Amodio and Devine (2006) is that the measures of implicit stereotyping and evaluation are in themselves behavioural expressions of bias, and not pure indices of an underlying process (Payne, 2005; Sherman, 2006). Therefore, participants' responses to the IATs could be determined by multiple processes. For example, although we predicted that scores on the evaluative IAT were primarily driven by affective

It is interesting to note that these findings are consistent with theorising about cognitive vs affective aspects of explicit self-reported attitudes (Dovidio et al., 1996; Dovidio, Esses, Beach, & Gaertner, 2004). Building on the theorising of Millar and Tesser (1986, 1989), Dovidio and colleagues proposed that affective measures of racial bias should predict consummatory responses, which involve interpersonal engagement, whereas cognitive measures should predict instrumental responses, which involve judgements, decisions, and impressions of outgroup members. This framework was generally supported by a meta-analysis of 31 studies (Dovidio et al., 2004). The findings of Amodio and Devine (2006) concerning the effects of implicit evaluation vs implicit stereotyping on behaviour dovetail nicely with the patterns of affective vs cognitive forms of bias observed by Dovidio et al. (2004). The multiple memory systems approach proposed by Amodio and Devine (2006) connects Dovidio et al.'s theorising to research on implicit bias and provides a model of the underlying neurocognitive mechanisms that give rise to these distinctions.

Neural correlates of implicit stereotyping vs evaluation. Although our research on the independence of implicit evaluation vs implicit stereotyping was grounded in neuroscience models, our initial tests of this model employed behavioural methods (Amodio & Devine, 2006). As a further step in establishing the validity of our theoretical framework, it was important to show that individual differences in implicit evaluation and stereotyping were uniquely associated with neural activity in regions associated with affective vs semantic processing. To this end, my colleagues and I recently conducted an experiment using fMRI, which allowed us to measure patterns of brain activity as participants engaged in judgements that involved either stereotyping or evaluative processing (Potanina, Pfiefer, Lieberman, & Amodio, 2008).

The fMRI scanning environment presents a unique set of challenges for social psychological experiments. Participants must lie very still on their

processes, it is plausible that the category judgements reflect semantic associations between the social groups (Black vs White) and general semantic concepts of positive and negative. In past research, participants' scores on an evaluative IAT were correlated with one's degree of amygdala activity in response to viewing Black compared with White faces (Cunningham et al., 2004a; Phelps et al., 2000; Potanina et al., 2008). However, other research has shown that a patient who acquired amygdala damage later in life showed an anti-Black/pro-White bias on an evaluative IAT (Phelps, Cannaistraci, & Cunningham, 2003). Considered together, these findings suggest that although implicit evaluative biases are typically driven by affective processes, similar associations may be made by semantic systems (i.e., by associating the concept "bad" with an outgroup member). An important remaining question, however, is whether the evaluative IAT score of an amygdala-damaged patient would be associated with changes in automatic arousal and non-verbal behaviours typically associated with implicit affective processes during interracial interactions.

backs while inside the bore of the scanner magnet, with their heads immobilised and their arms at their sides. Participants must view stimuli either through a mirror reflecting a monitor positioned behind the magnet or though LCD goggles, and make their responses by pressing buttons on a button box held in the right hand. Given these limitations, it is often desirable to create an engaging cover story and experimental task to counter that austerity and the constraints of the scanner environment (cf. Harmon-Jones, Amodio, & Zinner, 2007). For our purposes, it was important to engage participants in meaningful judgements that involved semantic or evaluative processing relevant to stereotyping and prejudice. By actively manipulating the task goals we were able to make stronger inferences about the meaning of any observed brain activity. By contrast, when participants passively view pictures without a specific task or goal, one cannot make very strong inferences about the meaning of their brain activations beyond basic aspects of visual processing.

Our study was introduced to participants as examining their ability to infer information about a target person based only on a picture of the person's face. Participants were told that the study was testing whether they could accurately infer a target person's preferences for certain activities, such as sports, or the likelihood that the target is the type of person with whom he or she would be friends. In line with the cover story, and to lead participants to believe that we could later assess the accuracy of their judgements, participants filled out questionnaires assessing their own preferences for various activities/hobbies and for qualities they preferred in a friend. They were then told they would make judgements 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 each pair was more likely to (a) be someone they would likely befriend (an evaluation-based judgement; Figure 4a) or (b) possess greater athletic abilities (a stereotype-based judgement; Figure 4b). 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). In addition, the pair of faces presented on each trial was always of the same race (Black, White, or Asian), and therefore judgements 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. Finally, once outside the scanner participants completed IATs measuring implicit evaluation and stereotyping, as in Amodio and Devine (2006).

Based on the neuroscience model of implicit race bias described above, we expected that participants scoring higher on the evaluative IAT would show

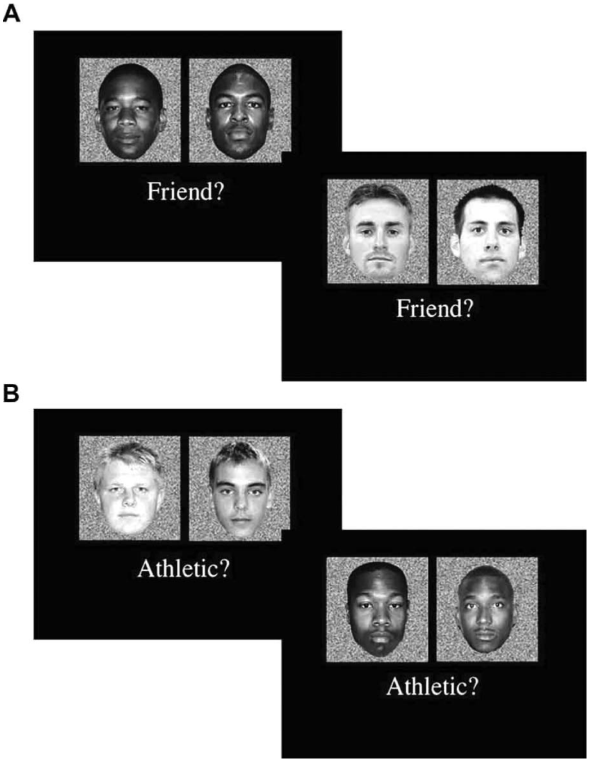

Figure 4. Stimuli for same-race judgements of faces, on the basis of perceived likelihood of friendship (A) or athletic ability (B), as used by Potanina et al. (2008).

greater amygdala activation when making friendship judgements of Black vs White faces, as suggested by past findings (Amodio et al., 2003; Cunningham et al., 2004a; Phelps et al., 2000). By contrast, we expected that participants scoring higher on the stereotyping IAT would show greater activation in the left posterior PFC—a region implicated in semantic processing—when making trait judgements of Black vs White faces. Consistent with these predictions, evaluative IAT scores, but not stereotyping IAT scores, were uniquely associated with amygdala activity during friendship judgements (but not trait judgements) (Figure 5A). By contrast, stereotyping IAT scores, but not evaluative IAT scores, were uniquely associated with left posterior PFC activity during trait judgements (but not friendship judgements) (Figure 5D).

In summary, the cumulative findings from this programme of behavioural and neuroimaging research provide discriminant and predictive validity for the multiple memory systems framework of implicit race bias.

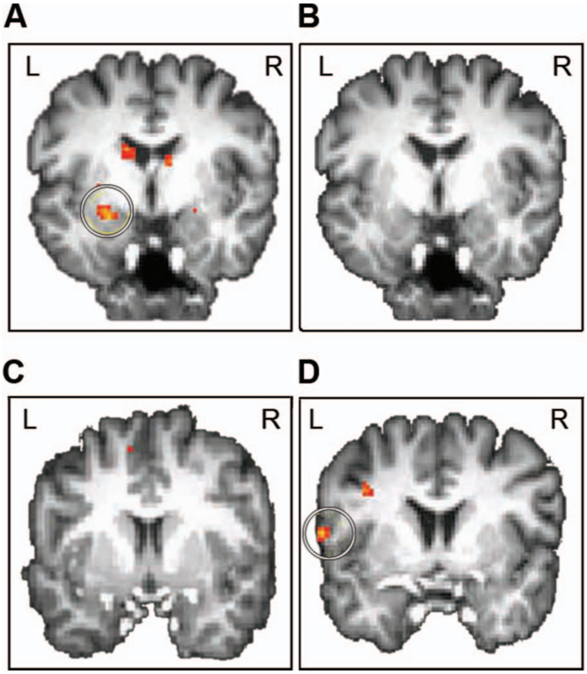

Figure 5. Correlations between IAT scores and neural activity in Potanina et al. (2008). During friendship judgements, stronger evaluative IAT scores were associated with greater amygdala activity (A), but stereotyping IAT scores were unrelated to amygdala activity (B). Neither IAT measure predicted activity in the left posterior prefrontal cortex (pPFC) during friendship judgements. During trait judgements, evaluative IAT scores were unrelated to pPFC activity (C), but larger stereotyping IAT scores predicted greater activity in the pPFC (D). During trait judgements, neither IAT measure predicted amygdala activity. To view this figure in colour, please see the online version.

Furthermore, these findings validated the notion that greater correspondence between measures of implicit bias and behaviour may be obtained when one considers the underlying neurocognitive mechanisms and their associated response channels (cf. Ajzen & Fishbein, 1977). In surveying past research on the correspondence between implicit measures and behaviour, our work suggests that the mixed findings in the literature may have to do with a lack of fit between implicit measures and behavioural outcomes (e.g., Fazio & Olson, 2003). According to our theoretical analysis, the lack of implicit attitude–behaviour correspondence observed in the literature may be due, in part, to models of implicit social cognition that do not distinguish between important sub-systems of learning and memory, and as a consequence, to the fact that research has not used measures or experiments that might reveal these critical distinctions.

Implications of social neuroscience model for intergroup anxiety

Up to this point I have described how a multiple-memory systems model of implicit bias, derived from the social neuroscience approach, serves to clarify past theorising on the construct of implicit bias and its link to discriminatory behaviour. Building on this theoretical advance, we have begun to explore the implications that this model may have for understanding the effects of intergroup anxiety associated with an interracial interaction. Several theorists have noted that concerns about prejudice during an interracial interaction often elicit anxiety and stress (Ickes, 1984; Mendes, Blascovich, Lickel, & Hunter, 2002; Shelton, 2003; Stephan & Stephan, 1985). For example, a White person interacting with an African American may monitor for unintentional expressions of racial stereotypes, whereas an African American may worry about behaving in ways that corroborate the African American stereotype (Gaertner & Dovidio, 1986; Steele & Aronson, 1995). Although several studies have investigated the effects of intergroup anxiety on mechanisms of control (e.g., Amodio, 2008; Lambert et al., 2003; Richeson & Shelton, 2003), few have examined how intergroup anxiety may modulate the activation and expression of implicit racial biases (e.g., Lambert et al., 2003).

In our theorising on the effect of anxiety on implicit bias, we considered the connections between the social neuroscience view of implicit stereotyping and evaluation and neuroscience models of anxiety. In the neuroscience literature, a large body of research links anxiety to activity of the amygdala and its related subcortical circuitry (Bishop, 2007; Cannistraro & Rauch, 2003). Given our findings that implicit evaluative bias also relates to amygdala activation but that implicit stereotyping does not, we hypothesised that engagement in an anxiety-eliciting interracial interaction should uniquely amplify the activation of implicit evaluative bias without directly affecting the activation of implicit stereotyping (Amodio & Hamilton, 2008). In a study testing this hypothesis, White participants were told they would be interacting with either a White or Black woman to discuss issues of social discrimination (but without specific reference to race). Participants were given an opportunity to jot down some potential discussion points on a sheet of paper prior to their interactions, as a way to engage them in thought about the upcoming discussion. Just before the interaction was to occur, participants completed an affect checklist, and then separate reaction-time measures of implicit evaluation and implicit stereotyping that were conceptually similar to the IATs used by Amodio and Devine (2006) but presented in the style of a flankers task (Payne, 2005). On each trial of the implicit evaluation measure, a picture of a White or Black male face was presented in the centre of the computer screen, and a pleasant or unpleasant word appeared simultaneously either above or below the picture. The participants' task was to ignore the picture and to categorise the word as good or bad. Trials on the implicit stereotyping task were similar except that target words consisted of intelligence- and athletic-related terms, which participants categorised as "mental" or "physical".

An initial set of analyses established that participants in both conditions exhibited significant levels of implicit stereotyping and anti-Black evaluation, and that participants in the interracial interaction condition reported greater anxious affect just prior to the interaction, compared with those in the same-race condition. These effects replicated past findings (Amodio & Devine, 2006; Lambert et al., 2003). Importantly, a set of analyses testing our specific hypotheses showed that participants who anticipated interacting with a Black person exhibited significantly higher levels of implicit evaluative bias compared with those in the same-race interaction condition, supporting our main prediction. By contrast, the race of the interaction partner did not moderate participants' levels of implicit stereotyping. Additional analyses, in which the process-dissociation procedure was used to estimate the independent contributions of automatic and controlled processes to responses on the implicit tasks (Jacoby, 1991; Payne, 2001), indicated that the anticipated interracial interaction was associated specifically with increased automatic evaluation. These findings begin to shed light on how implicit biases may operate in actual interactions and how they interact with emotional responses in such situations. More broadly, these results suggest that different aspects of implicit bias are expressed in different situations, with implicit evaluations playing a larger role in interpersonal interactions, whereas implicit stereotypes may play a larger role in decisions about outgroup members in the absence of personal contact.

Summary of social neuroscience approach to implicit race bias

To date, the social neuroscience approach has inspired much research on implicit race bias. Although initial research focused on associating implicit evaluation with amygdala activation, more recent research has begun to apply neuroscience models of learning and memory systems to enhance our understanding of the underlying mechanisms of implicit bias and their influence on behaviour. This programme of research has revealed an important distinction between affectively driven implicit evaluation and semantically driven implicit stereotyping that should serve to clarify long-standing questions about the nature of implicit racial biases and their expressions in behaviour. These two forms of bias map onto distinct memory systems, associated with classical conditioning and conceptual priming, respectively. On the basis of these neurocognitive associations, we predicted and found that implicit evaluations are more likely to be expressed in interpersonal behaviours (e.g., personal distance and non-verbals) and affected by intergroup anxiety, whereas implicit stereotypes are more likely to be expressed in judgements and impressions of outgroup members, without being affected by anxiety. Future research will build on these theoretical distinctions to consider the ways in which implicit evaluative and stereotypic associations may be regulated and ultimately unlearned.

THE SOCIAL NEUROSCIENCE OF PREJUDICE CONTROL

Once implicit racial biases become activated, what keeps them from influencing our behaviour? Most theorists agree that responding without prejudice often requires some form of self-regulation (e.g., controlled processing), whereby a person with non-prejudiced beliefs responds in an intended egalitarian manner despite the activation of biases that might lead to negative reactions or the use of stereotypes (Allport, 1954; Devine, 1989; Fazio, 1990). However, although the general idea that controlled processing is often required for responding without prejudice has been supported by a large body of research, critical questions remain about the nature of control, such as: What specific psychological process is being controlled? How do we know when control is needed? Is control a deliberative process or can it operate without awareness? An understanding of these basic aspects of control is key to understanding exactly how and under what conditions a person will be able to effectively implement an intended non-prejudiced response. For example, one of the most provocative findings in research on intergroup bias is that individuals with sincere egalitarian beliefs often show signs of implicit racial bias on behavioural and physiological measures (Amodio et al., 2003; Correll, Park, Judd, & Wittenbrink, 2002; Cunningham et al., 2004a; Devine, 1989; Devine et al., 1997, 2002; Greenwald et al., 1998; Monteith, 1993). Other research has shown that when under distraction (e.g., cognitive load), egalitarians tend to show signs of bias that do not appear in more deliberative responses (D. Gilbert & Hixon, 1991; Spencer, Fein, Wolfe, Fong, & Dunn, 1998). What explains these failures to regulate one's response to race? Why are some egalitarian individuals more effective at responding without prejudice than others? And does the control of bias operate differently when it is motivated by internal goals vs external social pressures? These are some unanswered questions about prejudice control that social neuroscience research has begun to address.

What does a social neuroscience approach bring to the study of prejudice control? Although social psychologists have grappled with questions about the nature of control for many years, a reliance on self-report and behavioural measures—the traditional tools of the trade—may have limited our ability to examine processes that cannot be self-reported or clearly observed in behaviour. That is, it may have focused our attention on more deliberative aspects of self-regulation while leaving less-deliberative aspects hidden from view. Recent research taking the social neuroscience approach has been able to advance our understanding of control by offering an expanded set of measures that permit the online assessment of control-related neural activity. In additional to its methodological offerings, this approach brings to bear neuroanatomical evidence that provides a ‘‘neural roadmap’’ for understanding how neural processes associated with control connect to neural processes associated with implicit stereotyping and evaluation, as well as with different behavioural response channels. In what follows, I address some of the critical questions posed above from a social neuroscience perspective, beginning with some basic definitional issues concerning the nature of ‘‘control.’’

What’s being controlled?

Whenever discussing issues of prejudice control, it’s worth asking: ‘‘What’s being controlled, anyway?’’ Indeed, the term ‘‘prejudice control’’ is frequently used to describe the process of how one responds without prejudice, yet its meaning is somewhat unclear. According to traditional social psychological models, control is typically conceived as the process of inhibiting or suppressing the activation of an unwanted bias. In other words, ‘‘prejudice control’’ refers to the processes of ‘‘turning down the volume’’ on automatically activated thoughts or feelings so that they do not influence one’s response. However, a review of the neuroscience literature indicates that regions of the frontal cortex activated during cognitive control tasks (e.g., the Stroop task), such as dorsal and ventral regions of the PFC, are primarily linked to structures that function to coordinate motivated behavioural responses (e.g., motor cortex, basal ganglia; Lehéricy et al., 2004) but have few, if any, direct connections to basic affective structures such as the amygdala (Gabbot, Warner, Jays, Salway, & Busby, 2005; Ghashghaei & Barbas, 2002).

A consideration of the neural architecture of these ‘‘control regions’’ suggests that mechanisms of control do not target the sources of bias in memory directly, but rather enhance the influence of intentional, goal-directed processes on behaviour such that the influence of activated biases is made irrelevant. Indeed, much research has documented people’s inability to successfully suppress unwanted thoughts or emotions (Gross & Levenson, 1993; Macrae, Bodenhausen, Milne, & Jetten, 1994; Monteith, Sherman, & Devine, 1998; Wegner, 1994), further suggesting that mechanisms of control are poorly suited for inhibiting sources of bias, but rather have their regulatory effects on behavioural output. This definition is a departure from the classic, Cartesian view of control as a process of suppressing unruly passions and intrusive thoughts, of which many contemporary views of control may be seen as a vestige. Returning to the issue of prejudice control: Findings from anatomical and functional neuroscience studies suggest that it is the behavioural expression of prejudice that is being controlled, rather than the source of bias. A truly egalitarian response is one that is unaffected by bias (e.g., Fiske & Neuberg, 1990), and thus *prejudice control* represents the ability to respond in an intentional (e.g., accurate) manner irrespective of the potentially biasing effects of automatic prejudices and stereotypes (Amodio, Devine, & Harmon-Jones, 2008; Payne, 2005).

How do we know when control is needed?

According to several social psychological models, effective self-regulation requires that one is aware of the potential for a response bias and able to correct for the bias as a response unfolds (e.g., Ajzen & Fishbein, 2000; Fazio, 1990; D. Gilbert, Pelham, & Krull, 1988; Wegener & Petty, 1997; Wilson & Brekke, 1994). This general view of control has provided a useful explanation for understanding deliberative forms of behaviour, but it does not account for how one initially detects the presence of bias. The question of how the need for control is detected has come to be referred to as the ‘‘homunculus’’ problem, because extant models of control in cognitive psychology and social psychology appear to assume that a ‘‘little man’’ (the homunculus) inside our heads just knows when control is needed and alerts us when bias is present. Although this issue has been largely ignored in theories of prejudice control, it has important implications for our understanding of how and under what conditions self-regulatory processes will be engaged and effectively implemented. For example, when a person fails to respond in an intentional manner, was it because he failed to detect that control was needed? Or was it because the need for control was detected, but efforts to implement an intentional response failed? A better understanding of the mechanisms of response control could shed light on questions of why some egalitarians are better at regulating their intergroup responses than others.

To address the ‘‘homunculus’’ problem of control, Botvinick, Braver, Barch, Carter, and Cohen (2001) proposed that there are separate cognitive processes for (a) determining when control is needed and (b) implementing the intended behaviour. In their model it is assumed that several different response tendencies are often simultaneously activated in the brain in response to internal and/or external cues. When two or more activated tendencies imply different behavioural responses, there is conflict in the system. The first component of Botvinick et al.’s (2001) model functions to monitor the degree of conflict present in the cognitive-motor system at any given moment, and thus is referred to as the *conflict-monitoring* process. As the degree of conflict rises, a second, *regulative* process is engaged to implement one's intentional responses, thereby overriding unwanted alternative tendencies. The regulative component requires cognitive resources and corresponds closely to the mechanism of control posited in most social psychological models (see Wegner, 1994, for a conceptually relevant theory of control). The neural substrates of these two components of control have been studied extensively using fMRI as well as event-related potentials (ERPs)—electrical activity from groups of cortical neurons firing in response to a discrete psychological event and measured using electroencephalography (EEG; Fabiani, Gratton, & Coles, 2000). Because ERPs are direct measures of electrical activity from firing neurons (vs blood flow as measured by fMRI), they can be measured in real-time with extremely high temporal resolution and thus provide unique information about the timing of neurocognitive processes. Across fMRI and ERP studies, conflict monitoring has been associated with activity of the dorsal anterior cingulate cortex (dACC; Figure 6) whereas the regulative process has been linked to activity in the dorsolateral (dl)PFC (see Figure 3; Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Carter et al., 1998; van Veen & Carter, 2002). Social neuroscientists interested in prejudice have applied the conflict-monitoring model to address mechanisms of prejudice control (Amodio et al., 2004, 2006; Amodio et al., 2008; Bartholow, Dickter, & Sestir, 2006). Whereas previous models of control posit that an unintended race-biased response reflects a failure to override implicit bias

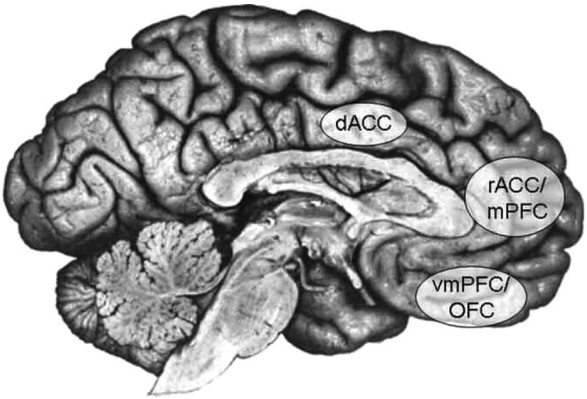

Figure 6. Medial view of the brain. Labelled regions include the dorsal anterior cingulate cortex (dACC), the general regions of the rostral anterior cingulate cortex (rACC) and medial prefrontal cortex (mPFC), and the general regions of the ventromedial prefrontal cortex (vmPFC) and orbital frontal cortex (OFC).

(corresponding to a failure of the regulative component; e.g., Devine, 1989), the Botvinick et al. (2001) model suggests that, alternatively, it might result from a lack of conflict arising in the response stream, such that a biased tendency is not recognised as conflicting with an intended non-biased response.

In an initial application of conflict-monitoring theory to questions of intergroup bias, my colleagues and I (Amodio et al., 2004) first sought to dissociate the process of conflict monitoring from response implementation in the context of stereotyping. To do this, we recorded participants' brain activity as they completed the weapons identification task (Payne, 2001; see below), from which we could derive specific ERPs that reflect the engagement of conflict-related ACC activity on different trial types. In each trial of the weapons identification task, a Black or White face prime was presented briefly (200 ms), followed by a target picture of either a handgun or handtool (Figure 7). Participants were instructed to categorise the target as a gun or tool irrespective of the prime. Responses on this task are primarily driven by stereotype associations of Black people as violent and dangerous, rather than by affective reactions (Judd, Blair, & Chapleau, 2004). Previous research has shown that the presentation of a Black face facilitates the identification of guns and interferes with the identification of tools (Payne, 2001). That is, Black faces activate a prepotent stereotypic association with guns, and participants often fail to inhibit this automatic tendency, such that they erroneously identify tools as ''guns'' after seeing a

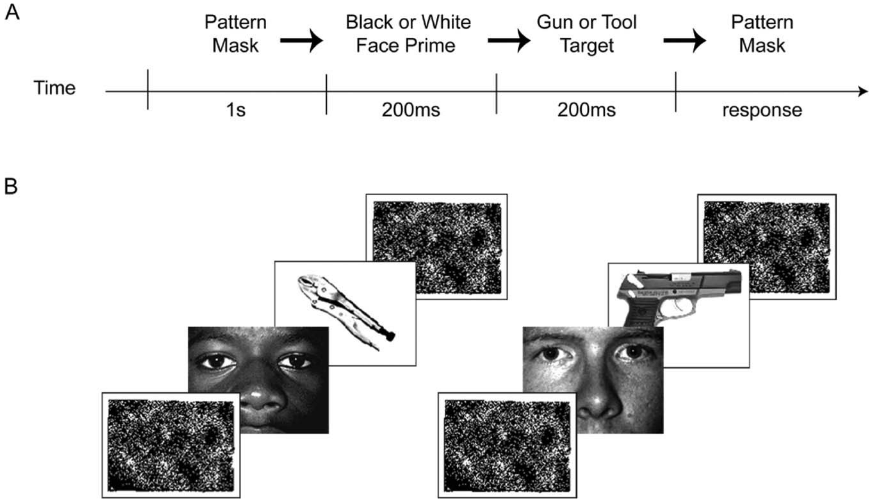

Figure 7. Schematic of weapons identification task, adapted from Payne (2001), illustrating the timecourse of events (A) and stimuli (B).

Black face. Our research focused on the role of the ACC in response control on this task. Past work has shown that a specific component of the ERP called the error-related negativity (ERN) indexes activity of the ACC that is related to conflict monitoring (Figure 8; Gehring, Goss, Coles, Meyer, & Donchin, 1993; van Veen & Carter, 2002; Yeung, Botvinick, & Cohen, 2004). Therefore, we measured the amplitude of the ERN wave when participants failed vs succeeded in responding without the automatic tendency to classify tools as guns following a Black face. By using an ERP measure of ACC activity, we could examine changes in neural activity on the order of milliseconds and thus study the timing of the conflict monitoring process as it unfolded during the course of a response.

Despite their motivation to respond without bias, participants in our study made a disproportionate number of errors on trials that required stereotype inhibition (i.e., responding ‘‘gun’’ on Black–tool trials). Nevertheless, when participants made this type of error, their ERN responses were larger than when they made errors on other types of trials, suggesting that their conflict-monitoring systems were detecting a heightened degree of response conflict, compared with other trials (Figure 9). When the intended (i.e., correct) response was congruent with the automatic tendency, such as when Black face primes were followed by pictures of guns, ERN amplitudes were relatively low. These findings demonstrated a dissociation between conflict-monitoring and regulatory aspects of control in the context of race bias, providing evidence that prejudice control is indeed a multi-component process, and that the detection of bias does not require deliberative processing (as suggested by previous social psychological models of control). The pattern of ERN responses was corroborated by another ERP wave associated with conflict monitoring, called the correct-related negativity

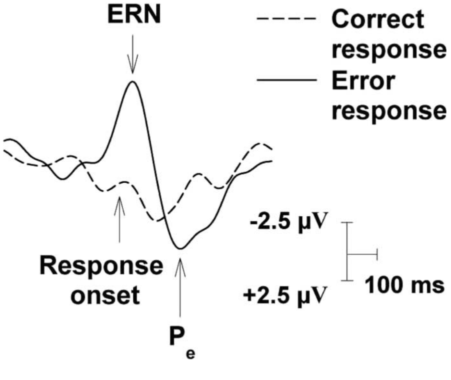

Figure 8. Illustration of the error-related negativity (ERN) and error positivity (P_e) components of an event-related potential (ERP).

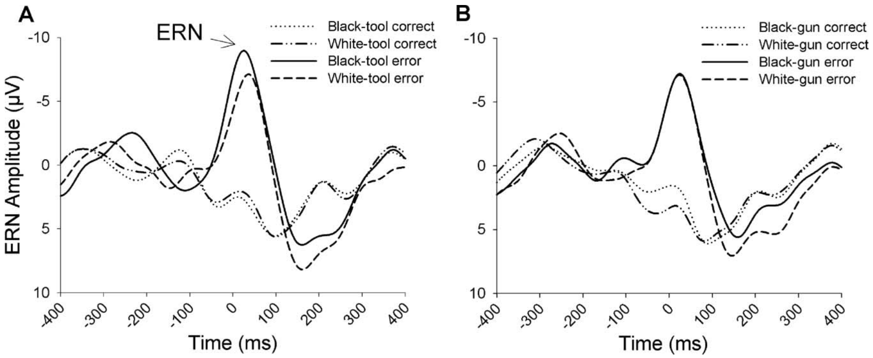

Figure 9. Response-locked event-related potential waveforms for correct and incorrect tool (A) and gun (B) trials as a function of race of face. The larger error-related negativity (ERN) elicited on Black–tool trials reflects the heightened activity of the conflict-monitoring system when an automatic stereotyping tendency conflicts with participants' intention to correctly categorise the target as "tool". Zero indicates the time of response.

(CRN; alternatively, the $N2_c$). The CRN occurs approximately 100–200 ms before a successfully controlled response and is generated by the same region of the ACC. These CRN results revealed that conflict-monitoring levels were also higher just prior to the successful control of automatic stereotype effects. Finally, we found that the magnitude of participants' ERN response on trials that required stereotype inhibition was strongly correlated with behavioural estimates of controlled processing (derived using the process-dissociation procedure; PDP, Jacoby, 1991; Payne, 2001), as well as behavioural accuracy on trials requiring stereotype inhibition. That is, participants with more sensitive conflict-monitoring systems were generally better at inhibiting stereotypes throughout the task. Taken together, these findings revealed that (a) the conflict associated with unwanted racial bias was detected independently of the implementation of control, and (b) individual differences in the sensitivity of this conflict-monitoring component of control strongly predicted the success of response control, thereby validating the utility of considering this component.

The role of conflict-related ACC activity in the regulation of stereotyped responses has since been replicated in subsequent ERP research (Amodio et al., 2006, 2008). Although fMRI studies have not yet shown a relationship between conflict-related ACC activity and the behavioural control of race bias, some research has shown that simply viewing faces of Black individuals elicits greater ACC and PFC activity, compared with viewing faces of White individuals (Cunningham et al., 2004a; Richeson et al., 2003). However, it is unclear whether some aspect of control is engaged in passive face-viewing paradigms. Therefore, additional research is needed to determine whether activations elicited by passively viewing faces might be related to controlled processing and the regulation of intergroup bias.

Explaining individual differences in egalitarians' ability to respond without bias

For the social psychologist, the primary appeal of the social neuroscience approach is that it promises to illuminate difficult social psychological questions from new angles. Having identified conflict monitoring as an important component in the regulation of prejudice, the next step was to apply the conflict-monitoring framework to address phenomena that have been difficult to explain with more traditional models of control. One such phenomenon is the finding that some egalitarian individuals have difficulty regulating their behavioural expressions of bias, compared with others who report equally egalitarian attitudes (Amodio et al., 2003; Devine et al., 1991, 2002; Monteith, 1993).

My colleagues and I hypothesised that variability in egalitarians' ability to inhibit expressions of automatic race bias may relate to the sensitivity of their conflict-monitoring systems. In our past work we found that differences in regulatory ability corresponded to people's internal and external motivations to respond without prejudice (Amodio et al., 2003; Devine et al., 2002). Internal motivation refers to personal reasons for responding without prejudice, whereas external motivation refers to normative reasons and worries about social disapproval, as measured by a questionnaire developed by Plant and Devine (1998). Individuals with egalitarian beliefs typically report being highly internally motivated to respond without bias, compared with those reporting low internal motivation. Interestingly, among the internally motivated egalitarians some report that they are also concerned about external social pressures and thus are motivated to respond without prejudice to avoid the disapproval of others. Drawing from self-determination theory (Deci & Ryan, 2000), Devine et al. (2002) suggested that being motivated by a combination of internal and external impetuses is characteristic of someone in transition to being more egalitarian (i.e., those in the process of "breaking the prejudice habit", Devine, 1989). These individuals reject prejudice in their consciously held beliefs, yet they have not yet internalised non-prejudiced responses into their dominant response set. As a result, they are effective in responding without bias in their deliberative behaviours, but show signs of bias in responses that are more spontaneous or when deliberation is unavailable (e.g., when under cognitive load). For convenience, I will hereafter refer to individuals fitting this profile as "poor regulators". By contrast, people who are internally motivated, but not concerned about external pressures, have internalised their non-prejudiced beliefs to the extent that responding without bias is autonomous and successfully implemented in both deliberative and spontaneous behaviours, irrespective of cognitive load (Devine et al., 2002). I will thus refer to these individual as ‘‘good regulators’’. Finally, individuals with low internal motivation to respond without prejudice are not expected to regulate either their deliberative or spontaneous responses to race (except for those with strong external motivations, when responding public). I refer to these individuals as ‘‘non-regulators’’.

We hypothesised that the previously observed differences in regulatory ability between good vs poor regulators may be due to differences in conflict-monitoring activity when a prepotent bias conflicts with an egalitarian response intention (Amodio et al., 2008). To test this hypothesis, we recruited participants fitting the three regulatory profiles described above on the basis of their scores on the Internal and External Motivations to Respond Without Prejudice scales (IMS/EMS; Plant & Devine, 1998), and recorded ERPs as they completed the weapons identification task. We focused on the responses of the good vs poor regulator groups. Both groups showed equivalent (and significant) levels of automatic stereotyping in their behaviour on the task (although these groups are known to differ in levels of implicit *evaluation*; Amodio et al., 2003; Devine et al., 2002). Both groups also reported positive explicit attitudes towards Black people (Brigham, 1993), and thus both needed to engage control in order to respond without stereotypes, in line with their explicit beliefs. However, good regulators exhibited greater controlled processing on the task, as indicated by PDP estimates, and responded more accurately on trials requiring the inhibition of stereotypes (i.e., Black–tool trials) than poor regulators, consistent with the notion that these individuals are more adept at regulating their intergroup responses. An examination of participants’ ERN responses indicated that differences in control between groups were associated with differences in conflict monitoring. Specifically, good regulators showed significantly larger ERN amplitudes than poor regulators on trials requiring the inhibition of automatic stereotypes, but did not differ on trials that did not require such control. Additional analyses showed that the group difference in response control on the task was mediated by participants’ ERN amplitudes. A third group characterised as non-regulators, which was included as a high-prejudice comparison, showed high levels of bias across measures and low levels of conflict monitoring. Collectively, these results suggest that the conflict monitoring mechanism for initiating control substantially accounts for the puzzling finding that some egalitarians are more effective in responding without bias than others.

Mechanisms for regulating bias according to internal vs external cues

A hallmark of social psychology is its emphasis on the power of the situation. For example, normative influences, such as pressure from peers or authority figures, can have profound effects on the ways people think and behave (Asch, 1956; Cialdini & Trost, 1998), and modern normative standards strongly proscribe expressions of racial bias (Crosby et al., 1980; Plant et al., 2003). Although traditional social psychological models do not distinguish between specific mechanisms underlying internal vs external forces on behaviour (cf. Carver & Scheier, 1978), several different lines of research suggest that internal and external impetuses for control may involve different processes. For example, in the intergroup literature, individual differences in the strength of personal and normative motivations to respond without prejudice tend to be independent (Dunton & Fazio, 1997; Plant & Devine, 1998). Research on motivation has identified different qualities of behaviour motivated by personal vs normative reasons, such that personally motivated behaviours tend to be more stable and consistent than normatively motivated behaviours (Deci & Ryan, 2000; Ryan & Connell, 1989). In light of these findings, it is possible that internally and externally driven forms of control may involve different underlying mechanisms that may relate to distinct neurocognitive processes. If this were the case, then a theory of self-regulation that accounts for these two different mechanisms would be more effective in predicting intergroup behaviour.

Interestingly, the notion that behaviours may be regulated by either internal or external impetuses for control has not been addressed by the neuroscience literature. However, recent neuroscience studies on empathy and mentalising are relevant to this issue because they concern the way an individual processes information about others (Frith & Frith, 1999). In neuroscience studies, empathy and mentalising are typically associated with activity in regions of the mPFC and rostral (r)ACC (see Figure 6; Harris & Fiske, 2006; Mitchell et al., 2005; Singer et al., 2004; for a review, see Amodio & Frith, 2006). Amodio and Frith (2006) noted that this region of mPFC lies at the intersection of neural regions associated with representations of one's own actions (the dACC) and representations of another person's actual and anticipated actions (anterior mPFC and orbital frontal cortex). Given the strong connectivity across this large region of cortex, we proposed that the mPFC supports a major regulatory function, whereby it is critical to the integration of representations of the self and others. Building on this theorising, my colleagues and I hypothesised that activations of the mPFC and rACC are important for externally driven forms of self-regulation, in contrast to dACC regions linked to the monitoring of internal regulatory cues (Amodio et al., 2006). We tested this hypothesis by measuring ERPs while participants completed the weapons identification task either (a) in private or (b) while being observed (via video monitor) by an experimenter for signs of prejudice. As in past work, the ERN component was taken as an index of conflict-monitoring processes. To assess activation of the rACC/mPFC, we examined the error-positivity (P_e) wave—a positive-polarity ERP component that immediately follows the ERN and is strongest at fronto-central scalp sites (see Figure 8; Hermann, Rommler, Ehlis, Heidrich, & Fallgatter, 2004; van Veen & Carter, 2002). In addition, whereas ERN responses have been shown to be independent of conscious awareness, the P_e is associated with the conscious perception of an unintended response (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). In testing our hypothesis, we preselected low-prejudice participants who reported being either high or low in sensitivity to external (normative) pressures to respond without prejudice, using Plant and Devine's (1998) scale. This way we could test the strong prediction that the P_e wave would be recruited for regulating race-biased behaviour only among highly externally motivated participants who responded in public.

As expected, larger ERNs amplitudes were associated with greater response control (i.e., less stereotype bias) across conditions and for all participants, consistent with the idea that the ERN reflects an internal cue to engage control that was present for all of the (low-prejudice) participants (Amodio et al., 2004). However, the P_e wave emerged as a strong predictor of control only among participants with high sensitivity to normative pressures who responded in public. Overall, this pattern of findings provided the first evidence that internally vs externally driven forms of prejudice control arise from different underlying neural mechanisms associated with the dACC and rACC/mPFC, respectively. These results suggest that externally motivated control may be less effective because it involves a more complex set of neurocognitive processes, compared with internally motivated control.

It is notable that this research provides another example of a hypothesis that was inspired by the convergence of social psychology and neuroscience. In this case, the notion of independent internal and external impetuses for control was first suggested by the social psychology literature. This hypothesis corresponded well to existing patterns of neuroscience data, and psychophysiological methods provided a way to test it. Finally, this study provides an example of how a social psychological idea can clarify neuroscientists' understanding of cortical function. The synthesis of social psychology and neuroscience in this work thereby provides a new set of theoretical ideas to be tested in future programmes of research.

Mechanisms for implementing intentional responses

The most important component of prejudice control is the implementation of an intended, non-biased response, because in the end, it is one's behaviour that causes discrimination. Much social neuroscience research on the control of intergroup responses has focused on the dorso- and ventro-lateral regions of the PFC (see Figure 3; Amodio et al., 2003, 2004; Bartholow et al., 2006; Cunningham, 2004; Lieberman et al., 2005, 2007; Richeson et al., 2003). Past work suggests that these lateral regions of the PFC are likely to support more intentional aspects of racial responses, on the basis of research in cognitive and affective neuroscience (Aron, Robbins, & Poldrack, 2004; Miller & Cohen, 2001; Rolls, 2000). Specifically, the dlPFC is generally believed to support the implementation of intentional (e.g., egalitarian) responses (Kerns et al., 2004), whereas the ventral lateral PFC (vlPFC) is believed to support the inhibition of unwanted tendencies (Aron et al., 2004).

In line with cognitive neuroscience theorising on PFC function, Richeson et al. (2003) found that exposure to Black (vs White) faces elicited activations of the ACC and dlPFC (but not the amygdala) in their White participants, which the authors interpreted as spontaneous efforts to exert control when viewing a Black face. The authors found that participants who showed greater dlPFC activity when passively viewing Black (vs White) faces later showed evidence of greater cognitive exertion during an interracial interaction. Similarly, participants in a study by Cunningham et al. (2004a) viewed faces of White and Black individuals and indicated whether the image appeared on the right or left side of the screen. When faces were presented supraliminally (525 ms; another condition presented faces for 30 ms), Black faces activated the ACC and dlPFC more than White faces. The authors interpreted these activations as evidence of prejudice control. Cunningham also found that activity in the vlPFC, believed to reflect inhibitory processes, was negatively correlated with amygdala activity elicited in the condition with 30 ms face presentations. Conceptually consistent effects have been reported by Lieberman et al. (2005, 2007). Given the fact that the vlPFC is not anatomically connected to the amygdala, Lieberman et al. (2007) proposed that the vlPFC-to-amygdala relationship is mediated by their respective connections to the mPFC. However, it remains unclear whether the vlPFC ‘‘down-regulates’’ the amygdala via the mPFC, or whether the mPFC is involved in modulating both the amygdala and regions of vlPFC.

Findings from fMRI studies that relate regions of PFC with responses to race are promising, in that they forge a link between research on prejudice control and more general models of control in cognitive neuroscience. However, additional research is needed to show that these activations are directly related to the control of prejudice. Indeed, few studies to date have reported a direct link between PFC activity and the regulation of bias and controlled patterns of behaviour (i.e., using a task that requires some form of response control). Furthermore, because the same regions of PFC implicated in prejudice control are also associated with many other psychological processes, including working memory, episodic retrieval, rehearsal, semantic monitoring, motivational orientation, and attentional gating (S. Gilbert et al., 2006), a strong experimental design and the use of behavioural measures to validate neural responses are needed for clearer interpretations.

Role of PFC in motivations associated with the regulation of intergroup bias

Recent research has begun to examine the motivational aspects of prejudice control, whereby control is associated with the motivated engagement of an egalitarian response (e.g., as opposed to the suppression of unwanted thoughts or feelings). For example, my colleagues and I recently used an electroencephalography (EEG) measure of dlPFC activity (cf. Pizzagalli, Sherwood, Henriques, & Davidson, 2005) to examine self-regulatory processes in the context of prejudice (Amodio et al., 2007). A large body of literature has suggested that left vs right asymmetries in PFC activity are associated with approach vs withdrawal motivation (Harmon-Jones, 2003), and we were interested in the roles of motivation and PFC activity in the regulation of race bias. We recruited a sample of participants who, on average, reported holding low-prejudice attitudes. Upon arrival, the participant was fitted with an EEG cap with embedded electrodes and baseline measures of EEG asymmetry and affect were assessed. Participants were then told they would view pictures of faces, followed by a set of pictures of objects and scenes. The block of face pictures consisted of 36 White, Black, and Asian faces displaying neutral expressions. The block of non-face pictures consisted of extremely positive (e.g., chocolate sundae), extremely negative (e.g., bloody corpse), and neutral (e.g., basket) pictures. After viewing these pictures, participants were told they would be shown graphs of their neural responses to the different types of pictures, via a computer program that would run automatically while the experimenter attended to other tasks. In fact, the feedback was bogus. First, participants were shown a graph indicating a more ‘‘positive’’ neural response towards positive pictures compared with neutral and negative pictures (with negative pictures eliciting the most negative neural response). This feedback was to be expected, and thus it served to bolster the perceived validity of the bogus results. Next, participants’ neural responses to the faces were displayed. These indicated that participants responded very positively to White faces, somewhat positively to Asian faces, and negatively towards Black faces. After viewing these bogus results, we measured participants’ baseline EEG and then their state affect. Compared with baseline, participants showed a decrease in left-frontal activity indicative of a reduction in approach motivation, as well as an increase in guilt (above and beyond changes in anxiety, sadness, anger towards others, or positive affect). Furthermore, decreased left-frontal cortical activity was correlated with increased guilt (but not any other emotion), indicating a coordination of motivational and emotional responses.

The roles of motivation, guilt, and PFC in the regulation of bias were tested in a second part of the experiment. Participants were told that the study was complete, but that in the 20 minutes remaining in the session we would like them to help us pretest stimuli for a future experiment. The future study was purported to involve reading various magazine articles, and we asked our participants to read the titles of these article and rate them according to how much one would personally be interested in reading each one. A subset of these articles pertained to reducing one's level of prejudice (e.g. "Improving your interracial interactions"), whereas others were unrelated (e.g., "Five steps to a healthier lifestyle"). The titles were presented one at a time, and participants were given 10 seconds in between each title to make their ratings. EEG was recorded while they viewed each of the titles. We found that participants reporting greater guilt in response to the bogus race-biased feedback in the first part of the study expressed greater desire to read articles about reducing prejudice, but guilt was unrelated to their desire to read articles unrelated to prejudice. In addition, a strong left-sided shift in left-frontal cortical activity (indicative of greater approach motivation) occurred as participants viewed these article titles, and the degree of left-side cortical activity was associated specifically with participants' desire to read the articles related to prejudice reduction. This study was unique in that it simultaneously tested basic questions about the interacting functions of emotion, motivation, and the frontal cortex and addressed a more specific question about the self-regulation of intergroup bias. Our results were consistent with the idea that mechanisms of self-regulation and motivation are engaged specifically to regulate behaviour, as opposed to thoughts or emotions per se, in line with previous theorising by Monteith (e.g., Monteith, 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Monteith & Mark, 2005). These results expanded on past research by showing the dynamic function of guilt and self-regulatory processes as one transitions from an initial reaction to the presence of bias to behaviour aimed at addressing one's bias.

Regulating bias by relating to others: An engagement hypothesis

Although current theories of the regulation of intergroup bias emphasise the suppression of unwanted biases and response tendencies, a growing body of findings is beginning to suggest an alternative mode of self-regulation that focuses on the engagement of self-other interactions. As described above, several studies in the field of cognitive neuroscience have linked regions of the medial frontal cortex with the processes of mentalising (also, Theory of Mind; Frith & Frith, 1999; Mitchell et al., 2005; Mitchell, Heatherton, & Macrae, 2002). Amodio and Frith (2006) suggested that this region functions to monitor the integration of one's own actions with the actions of others. Coupled with the findings of Amodio et al. (2006), in which the mPFC (and rACC) are more strongly engaged when regulating intergroup responses according to the presumed non-prejudiced norms of an observer, this interpretation of mPFC suggests an "engagement" model of self-regulation.

Although few studies have yet tested this model directly, findings from other areas of research appear to be consistent with this idea. For example, Harris and Fiske (2006) proposed that mPFC activity is associated with the process of humanisation—viewing another individual as possessing agentic and uniquely human properties (Haslam, 2006). Prior theorising by Fiske, Cuddy, Glick, and Xu (2002) related humanisation to two major factors of person perception—warmth and competence—and suggested that individuals who are perceived as low on both dimensions are seen as less-than-human, or *dehumanised*. Analyses of participants' warmth and competence ratings of a host of different social groups showed that among these, African Americans were perceived as lower on both dimensions than White Americans. That is, compared with White people, Black people were seen as less human. In an effort to link humanisation to social neuroscience studies of mentalising, Harris and Fiske (2006) had participants view pictures of people from groups classified as humanised (e.g. middle-class Americans) and dehumanised (e.g., homeless people), on the basis of Fiske et al.'s (2002) model. As expected, a comparison of brain activity for humanised vs dehumanised targets yielded significant activation in the same regions of mPFC linked previously to mentalising. That is, humanisation was related to greater mPFC activity, and by association, to the process of mentalising. This interpretation was corroborated by the authors' finding that, for dehumanised targets, lower mPFC activity was associated with ratings of lower feelings warmth (Harris & Fiske, 2008). An implication of these findings is that the regulation of intergroup responses may involve a process of "re-humanisation", whereby formerly dehumanised individuals come to be viewed as worthy of mentalising and social engagement. According to this analysis of re-humanisation as a means for responding without bias, regulation does not involve the suppression of unwanted thoughts and feelings, but rather the process of seeing a person in a new light and adopting a motivation to socially engage. Although Harris and Fiske (2006) did not examine perceptions of different racial groups, it is possible that their findings regarding the role of the mPFC in re-humanisation would generalise to White Americans' perceptions of African Americans.

More recently, research in my lab has used fMRI to test the possibility that the regulation of affective responses to outgroup members involves activity of the mPFC, which is linked to mentalising, self–other integration, and humanisation, rather than regions of lPFC associated with top-down suppression of unwanted bias. On the other hand, we theorised that a more traditional form of control involving top-down inhibition of unwanted thoughts would be recruited when participants attempted to inhibit the influence of stereotypes on their judgements, which should involve activation of the lPFC but not the mPFC. We adapted the fMRI scanner task used by Potanina et al. (2008), in which participants judged pairs of faces on the basis of which is more likely to be one's friend (an affect-related judgement) or likely to be interested in athletic activities (a stereotype-related judgement). In this version, participants made judgements of same-race pairs (i.e., pairs of White, Black, and Asian faces) as well as mixed-race pairs. The critical difference was that participants' concerns about responding with bias are relevant when making mixed-race judgements, but not on same-race trials. In line with this assumption, participants took longer to make judgements of mixed-race pairs than same-race pairs, and brain regions associated with self-regulation (e.g., ACC, mPFC, dorsolateral PFC) were more strongly activated during mixed-race judgements compared with same-race judgements. The primary analysis in this study examined contrasts of brain activity for friendship vs trait judgements of mixed-races. We found that mixed-race judgements about friendship uniquely elicited greater activity in the mPFC, as in studies of mentalising and humanisation, but not the vlPFC, consistent with the idea that these forms of socio-cognitive engagement represent a form of self-regulation (Figure 10A). By contrast, stereotype judgements were uniquely associated with activity in the vlPFC, an area linked to top-down inhibitory processes in past work (Figure 10D), but were not associated with the mPFC. Hence, this research suggests that the regulation of affective (and evaluative) responses to race may be more accurately characterised as a process of engagement, whereas the regulation of stereotypic beliefs may correspond better to traditional top-down models of control. This theoretical analysis was uniquely inspired by the convergence of neuroscience and social psychological findings, and tested by integrating methods from both fields.

Summary of social neuroscience contributions to prejudice control

Social neuroscience research on the topic of prejudice has revealed that the self-regulation of intergroup responses involves multiple coordinated underlying processes. These are summarised in Table 2. An understanding of the multi-process nature of prejudice control is important because it

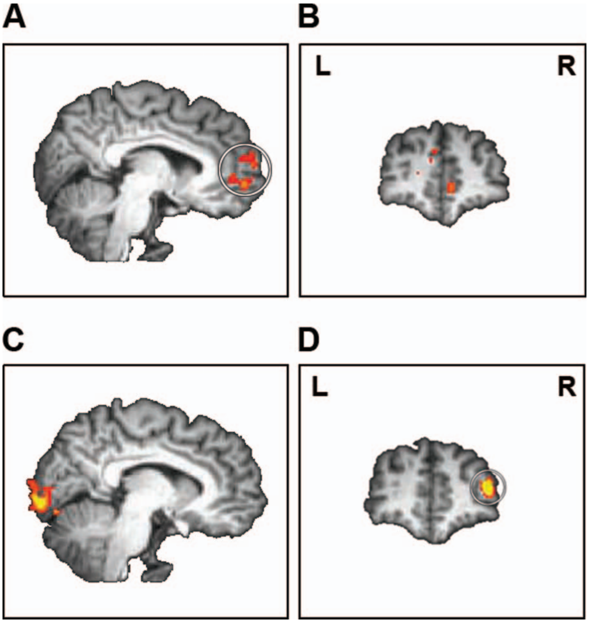

Figure 10. Patterns of neural activation associated with regulating friendship vs trait judgements about mixed-race (Black vs White) face pairs in Amodio and Potanina (2008). Judgements of whether one would be more likely to befriend Black or White person elicited greater medial prefrontal cortical activity (mPFC; panel A), but not ventrolateral prefrontal cortical (vlPFC) activity (panel B), compared with judging which person was more athletic. By contrast, judging which person was more athletic was unrelated to mPFC activity (panel C), but enhanced activity in the vlPFC (panel D), compared with judgements of potential friendship. To view this figure in colour, please see the online version.

TABLE 2
Processes involved in intergroup bias and their associated neurocognitive functions and neural correlates

Role in intergroup bias	*Neurocognitive function*	*Candidate structure(s)*
Implicit evaluative bias	Classical fear conditioning; arousal; vigilance	Amygdala
Implicit stereotyping	Semantic retrieval/selection, Conceptual priming	Left posterior PFC
Detecting bias/internal cues for regulation	Conflict monitoring	Dorsal ACC
Regulation according to social cues; engagement	Mentalising; integrating actions of self and other	Medial PFC, rostral ACC
Inhibition of stereotypes	Response inhibition	Right ventrolateral PFC
Implementation of an intended response	Regulative control	Anterior dorsolateral PFC

PFC = prefrontal cortex; ACC = anterior cingulate cortex.

suggests that failures to respond without prejudice may involve impairments at several different stages of control, which may result in different forms of biased behaviour. Additionally, different aspects of control may be differentially affected by situational factors (e.g., distraction, external social pressures). Finally, the identification of distinct underlying mechanisms is important because it suggests that individual differences likely exist in each of these processes. Although this review has focused on the role of self-regulatory processes in the context of intergroup bias, these are domain-general mechanisms that function to regulate a broad range of responses.

A SOCIAL NEUROSCIENCE PERSPECTIVE ON MECHANISMS OF PREJUDICE REDUCTION

A major goal of research on intergroup bias is to find ways to reduce prejudice in society. Most contemporary approaches to prejudice reduction in the field of social psychology have focused on changing prejudiced attitudes and reducing the strength of race-biased associations in implicit memory. Indeed, the ultimate goal of a prejudice reduction intervention is to rid a person of all implicit and explicit racial biases. But prejudice may be reduced in others ways as well, such as through the enhancement of self-regulation (Monteith et al., 2002) and through the modulation of situational factors (Hewstone, 1996). In the end, the goal of prejudice reduction efforts is to eliminate the expression of discriminatory behaviour, and thus any strategies that accomplish this goal are worth considering. Advances in the understanding of intergroup bias from the social neuroscience approach have pertained primarily to basic questions of automaticity and control in the context of intergroup responses, and thus this approach has the most potential to contribute to research on reduction at the intra- and inter-personal levels of analysis. At this time it is not clear whether neuroscience approaches will offer new insights into prejudice reduction as it occurs at the group and societal levels of intergroup relations. In this section I discuss some implications of social neuroscience research for different means of prejudice reduction.

Unlearning implicit racial associations

Since the discovery of robust implicit racial biases, prejudice researchers have focused much of their attention on strategies designed to weaken implicit racial associations. Yet despite the attention given to this endeavour, there are relatively few examples of successful implicit race bias reduction reported in the literature (with demonstrations of malleability effects notwithstanding; e.g., Dasgupta & Greenwald, 2001; Kawakami, Dovidion, Moll, Hermsen, & Russin, 2000; Kawakami, Phills, Steele, & Dovidio, 2007; Olson & Fazio, 2006; Rudman, Ashmore, & Gary, 2001).

I have suggested that efforts to reduce implicit bias could be facilitated by considering the operating parameters of the memory systems that underlie implicit stereotyping and evaluation (e.g., Amodio, in press; Amodio & Devine, 2006). Extant research on implicit bias reduction has generally assumed that all forms of implicit bias, including stereotypes and evaluations, and attitudes more generally, are learned and stored in the same type of semantic network (Fazio et al., 1995; Greenwald & Banaji, 1995). Given theorising on associative learning systems (Sloman, 1996; Smith & DeCoster, 2000), researchers have assumed that implicit biases can be unlearned through repeated exposure to bias-inconsistent stimuli (e.g., Rydell & McConnell, 2006). This view comports with the multiple-memory systems framework regarding semantically based forms of implicit bias, such as associations with stereotypic content and good/bad concepts, which are presumably learned and unlearned in a slow, probabilistic fashion. However, this view does not correspond well with research on affective systems, which are known to learn and unlearn in a different way. Research on classical fear conditioning has shown that strong and enduring affective associations may be acquired from a single experience and that such associations are highly resistant to change. After a long period of non-activation, classically conditioned associated are easily "reconditioned", suggesting that although such associations may lie dormant, they may never be truly eradicated (see also LaBar & Phelps, 2005). That is, although the expression of a classically conditioned effect may habituate, the amygdala (specifically, cells in the lateral nucleus) will continue to respond to the CS (Maren & Quirk, 2004). Thus, the reduction of implicit stereotyping and evaluation may require different strategies, bringing with them unique sets of challenges.

The social neuroscience analysis of implicit bias posits that strategies for reducing implicit stereotyping may be somewhat different from those used for reducing implicit evaluation and, likewise, that changes in either form of bias may only appear in responses on tasks designed to assess that particular form of bias. For example, techniques for reducing implicit stereotyping should focus on repeated pairings of a stigmatised target with stereotype-inconsistent (or irrelevant) semantic stimuli (e.g., with words rather than pictures). We would expect that changes in implicit stereotyping would be most evident on implicit tasks assessing conceptual associations between stigmatised group members, but would not be as evident in measures of affective associations. By contrast, attempts to alter implicit evaluations should focus on decoupling more gut-level affective responses from images of stigmatised group members. Because the amygdala and its associated subcortical circuitry are not known to process semantic information (e.g., language), implicit evaluation reduction may be accomplished most effectively by the use of affectively laden imagery and experiences rather than positive or negative words. Moreover, changes in evaluative associations are likely to be revealed on affect-based measures of bias rather than stereotype-based measures (cf. Dovidio et al., 1996). However, given research suggesting that classical conditioning is not very amenable to extinction, one might expect lower success in the unlearning of implicit evaluation compared with implicit stereotypes.

Although there continues to be relatively little evidence for successful implicit bias reduction, the successful examples are notable because they appear to conform to the multiple memory systems model. For example, Kawakami et al. (2000, see also Kawakami, Dovidio, & van Kamp, 2005) focused on reducing the strength of participants' implicit stereotypes. The authors used an intervention in which participants completed several hundred trials of responding to stereotype-inconsistent word pairings, and changes in bias were assessed using a reaction-time measure of associations between stereotype-related words and either word labels of target groups or supraliminal pictures of target group members that could be semantically processed. That is, manipulations and measures involved semantic memory systems and abided by the parameters of semantic associative learning.

Other research has used affect-based strategies of bias reduction and has assessed changes in bias using measures of implicit evaluations (Dasgupta & Greenwald, 2001; Olson & Fazio, 2006). For example, participants in Olson and Fazio's (2006) studies viewed a series of positive and negative images and words. In the critical condition, pictures of Black individuals were presented in conjunction with positive images, and images of White individuals were presented with negative images. In a second condition, these pairings were reversed. Results showed that participants in the positive–Black/negative–White condition showed less of a racial bias on an evaluative association task both immediately afterwards and a day later. In a conceptually similar set of studies, Kawamaki et al. (2007) tested a strategy in which participants made approach or avoidance movements while being exposed to faces of White or Black individuals, and found that participants trained to approach Black people (and avoid White people) showed a reduction in implicit evaluative bias, as well as less discomfort during an actual interracial interaction. These findings are in line with the notion that evaluative bias reflects activity of basic affective systems (i.e., the amygdala and related subcortical structures) that function to orchestrate basic approach/avoidance responses (Lang et al., 1990; LeDoux, 1996). To date, research has not directly tested the idea of a double dissociation in reduction effects, such that semantic-based interventions should be primarily effective for changing implicit stereotyping, whereas affective-based interventions should primary alter implicit evaluations, as suggested by Amodio and Devine (2006).

Despite some successful instances of implicit bias reduction, this approach to reducing prejudice faces several challenges. First, implicit biases are believed to reflect a lifetime of exposure to racially biased information in one's cultural environment (Devine, 1989; Rudman, 2002). Exposure to a few hundred trials of bias-inconsistent stimuli in experimental tasks may be little match for years of enculturation. Furthermore, as soon as a participant leaves the experimental session, he or she is re-exposed to culturally embedded biases that work to undo any lab-based reductions in implicit racial associations. Although some research has shown that the effects of a particular intervention may last a day or more (Dasgupta & Greenwald, 2001; Olson & Fazio, 2006), questions of the practicality of such interventions over the long term remain. Finally, a major limitation of research on implicit bias reduction concerns the methodological challenges of assessing change at the implicit level of processing. That is, it is difficult to determine whether a change on a behavioural measure of implicit bias, such as the IAT, reflects (a) the unlearning of race-biased associations, (b) a temporary change in the accessibility of bias, or (c) an enhancement of controlled processing. In light of evidence that virtually all behavioural responses, including those made on reaction-time measures of bias, reflect some combination of automatic and controlled processing (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Jacoby, 1991; Payne 2001, 2005), it may be impossible to determine whether an implicit association has truly changed on the basis of behavioural assessments. On the bright side, however, it may be argued that while knowing whether an implicit association has changed is an important theoretical issue, it is not necessarily a practical issue given that the larger goal of reducing prejudice is to abate the behavioural expressions of discrimination regardless of whether an underlying racial association exists.

Reducing expressions of implicit bias by enhancing control

Early research on the control of intergroup responses assumed that controlled processes required a great deal of effort and deliberation (Devine, 1989; D. Gilbert & Hixon, 1991). For this reason, it did not seem practical to focus on enhancing people's ability to control bias, given that deliberative controlled processing would not be engaged in spontaneous responses, such as during the rapid exchange of a lively interpersonal interaction. However, this view of control as a very deliberative process has changed substantially over the past decade, with several studies demonstrating that aspects of response control may be engaged with little or no conscious deliberation (Amodio et al., 2004, 2008; Mendoza, Gollwitzer, & Amodio, 2008; Monteith et al., 2002; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). For example, the conflict-monitoring component of control arises from competition among implicit response tendencies and has been shown to operate without conscious awareness (Botvinick et al., 2001; Neiuwenhuis et al., 2001). My colleagues and I have demonstrated the role of conflict monitoring in the regulation of automatic stereotyping (Amodio et al., 2004) and have suggested that interventions designed to enhance the sensitivity of this system to cues of bias may provide an effective strategy for non-effortful prejudice reduction (Amodio et al., 2008). Similarly, a programme of research by Monteith and her colleagues (Monteith, 1993; Monteith et al., 2002; Monteith & Mark, 2005) demonstrates that training people to recognise cues for potential race bias is effective in reducing behavioural expressions of race bias.

Recently my colleagues and I have explored the use of implementation intentions as a non-deliberative strategy for reducing expressions of implicit race bias (Mendoza et al., 2008). Implementation intentions are if-then plans that link a goal-directed response to a specific situational cue, such as "If I see a Black person, I will respond more carefully." The purpose of this type of implementation intention is to ensure that intended response is engaged spontaneously when the potential for bias becomes present. In our study participants completed the Shooter Task, in which Black and White males appear in different background scenes holding either a gun or an innocuous object. Participants are instructed to "shoot" armed targets but not unarmed targets, by pressing buttons labelled "shoot" or "don't shoot" on the computer keyboard. A pattern of race-biased responding is typically elicited by this task, such that unarmed targets are erroneously "shot" more often if they are Black than if they are White. Participants provided with the implementation intention ("If I see a gun I will shoot" and "if I see an object I will not shoot") for completing the task showed significantly lower levels of implicit bias on the task than participants receiving a simple goal strategy ("I will always shoot a person I see with a gun!" and "I will never shoot a person I see with an object!") or participants in a control condition who did not receive a strategy. Furthermore, process-dissociation analyses determined that implementation intentions enhanced controlled patterns of responding, but did not affect automatic processes, relative to the other conditions. These initial results indicate that methods that facilitate the engagement of control may provide a practical strategy for reducing more spontaneous expressions of implicit race bias.

Controlling situational influences

Most discussions of prejudice reduction focus on eliminating racist attitudes and beliefs. However, an often-underappreciated mode of prejudice reduction involves the reconfiguration of situational factors in a way that reduces bias and promotes intergroup harmony. One widely studied situational intervention is intergroup contact (Allport, 1954; Brown &

Hewstone, 2005). Decades of research on intergroup contact has shown that, under the right conditions, personal contact among members of different groups will lead to reduced expressions of prejudice (Hewstone, 1996; Pettigrew, 1998). However, several theorists have noted that anxiety associated with an interracial interaction may undermine the prejudice-reducing effects of contact (Richeson & Shelton, 2003; Stephan & Stephan, 1985). Our recent research linking a generalised amygdala-based substrate of anxiety to the amplification of implicit evaluation (Amodio & Hamilton, 2008) suggests that general anxiety-reducing procedures, even if they have little to do with a particular intergroup dynamic, should facilitate the effects of contact by reducing the activation of implicit evaluative bias. That is, intergroup contact procedures should be more successful when they take place in an anxiety-reducing context.

Along similar lines, much recent research has shown that the context in which one perceives an outgroup member affects the degree to which automatic intergroup biases are activated (Barden, Maddux, Petty, & Brewer, 2004; Lowery, Hardin, & Sinclair, 2001; Wittenbrink et al., 1997). For example, a picture of a Black person is less likely to elicit automatic stereotyping when it is superimposed over a non-threatening background (e.g., a church) versus a threatening background (e.g., a dark alley; Wittenbrink et al., 1997). Interpretations of such findings have focused on the malleability of individuals' representations of African Americans. However, these effects suggest another form of prejudice reduction that focuses on manipulating the situational environment in such a way as to attenuate the activation of implicit biases. These effects may be driven in part by the non-deliberative forms of control and interpersonal engagement associated with ACC and mPFC functions—a hypothesis suggested by social neuroscience perspectives on self-regulation.

Summary of social neuroscience implications for prejudice reduction

In summary, the social neuroscience approach suggests some refinements to existing models of implicit prejudice reduction. Drawing on the multiple-systems model of learning and memory, this approach posits that implicit stereotyping and implicit evaluation may be extinguished through different processes and therefore should be targeted by different reduction techniques. Given the finding that implicit evaluation and stereotyping may be expressed through different response channels, it is likely that interventions targeting affective associations would primarily reduce bias in interpersonal behaviours, whereas those targeting semantic associations would primarily reduce bias in higher-level processes involving judgements and impression formation (Amodio & Devine, 2006; Dovidio et al., 2004). It will be important for future research to include a broader set of outcome measures of expressions of bias in order to capture the full range of effects that reflect intervention-based changes in different underlying memory systems. Furthermore, the theoretical analysis inspired by the social neuroscience approach suggests that although it may be very difficult to extinguish race-biased associations, spontaneous forms of controlled processing and attention to situational factors may constitute more effective means of prejudice reduction. The broader take-home message of this review is that the most effective programme of prejudice reduction should involve a multi-pronged strategy that targets each of the subcomponents of implicit and explicit biases independently and also takes specific situational moderators of bias into account. It is equally important to evaluate the effects of such interventions using a broad range of outcome measures.

CONCLUSION

The strength of the social neuroscience approach to intergroup relations lies in its ability to shed new light on the basic psychological mechanisms involved in stereotyping and prejudice by integrating models of learning, memory, and emotion from the neuroscience literature. Through such integrations, this nascent but rapidly developing approach has already begun to yield new insights into some of the field's most difficult questions. As described in this review, this approach has begun to clarify the basic constructs of implicit bias and the pathways through which they relate to behaviour. The social neuroscience perspective has also begun to show how mechanisms of self-regulation may be engaged in a relatively spontaneous fashion, in addition to the more deliberative modes of control studied previously in the social psychological literature. Finally, a consideration of the neurocognitive research on learning and memory provides new predictions for how racial associations may be unlearned. An important implication of this analysis is that unlearning implicit biases may be difficult, but that efforts to enhance spontaneous forms of control and to modify environmental factors may prove to be a more effective strategy for prejudice reduction. As research on intergroup relations continues to evolve, I expect that the social neuroscience approach will continue to serve the goals of understanding and reducing prejudice by providing new insights into how the basic elements of intergroup perceptions, emotions, and behaviours interact and relate to general mechanisms of the mind.

REFERENCES

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

Ajzen, I., & Fishbein, M. (1977). Attitude–behaviour relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude–behaviour relation: Reasoned and automatic processes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 1–33). Chichester, UK: John Wiley & Sons.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Amodio, D. M. (2008). *Intergroup anxiety effects on the regulation of stereotypes: A psychoneuroendocrine analysis*. Manuscript submitted for publication.

Amodio, D. M. (in press). Self-regulation in intergroup relations: A social neuroscience framework. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.) *Social neuroscience: Toward understanding the underpinnings of the social mind*. New York: Oxford University Press.

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

Amodio, D. M. & Devine, P. G. (in press). On the functions of implicit prejudice and stereotyping: Insights from social neuroscience. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures*. New York: Psychology Press.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, *18*, 524–530.

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

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

Amodio, D. M., Hamilton, H. K., & Potanina, P. V. (2008). *Interracial interaction anxiety amplifies implicit race-biased evaluations but not implicit stereotyping*. Unpublished manuscript.

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 responses 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., & Lieberman, M. D. (in press). Pictures in our heads: Contributions of fMRI to the study of prejudice and stereotyping. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Amodio, D. M., & Potanina, P. V. (2008). *Regulating social responses through inhibition vs. self-engagement: An fMRI study*. Unpublished manuscript.

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

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychological Monographs, 70(9).

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgement. *Social Cognition*, *21*, 61–87.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22.

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.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*, 563–573.

Bechara, A., Damasio, H., & Damasio, A. R. (1995). Role of the amygdala in decision-making. *Annual Review of Neuroscience*, *985*, 356–369.

Bishop, S. J. (2007). Neurocognitive mechanisms of anxiety: An integrative account. *Trends in Cognitive Science*, *11*, 307–316.

Blair, I. (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: Lawrence Erlbaum Associates Inc.

Blaxton, T. A., Bookheimer, S. Y., Zeffiro, T. A., Figlozzi, C. M., William, D. D., & Theodore, W. H. (1996). Functional mapping of human memory using PET: Comparisons of conceptual and perceptual tasks. *Canadian Journal of Experimental Psychology*, *50*, 42–56.

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

Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179–181.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behaviour Processes*, *20*, 219–231.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied and Social Psychology*, *23*, 1933–1967.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 255–343). San Diego, CA: Elsevier.

Cacioppo, J. T., Berntson, G. G., Lorig, T. S., Norris, C. J., Rickett, E., & Nusbaum, H. (2003). Just because you're imaging the brain doesn't mean you can stop using your head: A primer and set of first principles. *Journal of Personality and Social Psychology*, *85*, 650–661.

Cannistraro, P. A., & Rauch, S. L. (2003). Neural circuitry of anxiety: Evidence from structural and functional neuroimaging studies. *Psychopharmacological Bulletin*, *37*, 8–25.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.

Carver, C. S., & Scheier, M. F. (1978). Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, *36*, 324–332.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 151–192). New York: McGraw-Hill.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

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.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*, 546–563.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004a). 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. (2004b). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*, 1717–1729.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Science*, *3*, 11–21.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, *15*, 353–375.

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

Deci, E. L., & Ryan, R. M. (2000). The ''what'' and ''why'' of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry*, *11*, 227–268.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. E. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, *15*, 5870–5878.

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.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*, 835–848.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice and discrimination: Another look. In C. N. McCrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). New York: Guilford Press.

Dovidio, J. F., Esses, V. M., Beach, K. R., & Gaertner, S. L. (2004). The role of affect in determining intergroup behaviour: The case of willingness to engage in intergroup affect. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 153–171). Philadelphia: Psychology Press.

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., 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.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist*, *60*, 181–190.

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 53–84). New York: Cambridge University Press.

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

Fazio, R. H. (1990). Multiple processes by which attitudes guide behaviour: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). New York: Academic Press.

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude–behaviour consistency and the strength of the object–evaluation association. *Journal of Experimental Social Psychology*, *18*, 339–357.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297–327.

Fendt, M., & Fanselow, M. S. (1999). The neuroanatomical and neurochemical basis of conditioned fear. *Neuroscience and Biobehavioural Review*, *23*, 743–760.

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.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation: Form category-based to individuating process as a function of information, motivation, and attention. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–108). San Diego, CA: Academic Press.

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

Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centres. *Journal of Comparative Neurology*, *492*, 145–177.

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

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). San Diego, CA: Academic Press.

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

Gale, G. D., Anagnostaras, S. G., Godsil, B. P., Mitchell, S., Nozawa, T., Sage, J. R., et al. (2004). Role of the basolateral amygdala in the storage of fear memories across the adult lifetime of rats. *Journal of Neuroscience*, *24*, 3810–3815.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Gazzaniga, M. S. (2004). *The cognitive neurosciences III*. Cambridge, MA: MIT Press.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, *115*, 1261–1279.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509–517.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*, 733–740.

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.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition. *Psychological Review*, *102*, 4–27.

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

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behaviour. *Journal of Personality and Social Psychology*, *64*, 970–986.

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology*, *40*, 838–848.

Harmon-Jones, E., Amodio, D. M., & Zinner, L. R. (2007). Social psychological methods in emotion elicitation. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 91–105), NewYork: Oxford Universtiy Press.

Harris, L. T., & Fiske, S. T. (2006). Dehumanising the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*, 847–853.

Harris, L. T., & Fiske, S. T. (2008). *Disgust evokes less mentalising and avoidance of social targets*. Unpublished manuscript.

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.

Herrmann, M. J., Rommler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/N_e) and positivity (P_e). *Cognitive Brain Research*, *20*, 294–299.

Hewstone, M. (1996). Contact and categorisation: Social psychological interventions to change intergroup relations. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 323–368). New York: Guilford Press.

Haslam, N. (2006). Dehumanisation: An integrative review. *Personality and Social Psychology Review*, *10*, 252–264.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Associates Inc.

Ickes, W. (1984). Compositions in black and white: Determinants of interaction in interracial dyads. *Journal of Personality and Social Psychology*, *47*, 330–341.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Judd, C. M., Blair, I. V., & Chapleau, K. M. (2004). Automatic stereotypes versus automatic prejudice: Sorting out the possibilities in the Payne (2001) weapon paradigm. *Journal of Experimental Social Psychology*, *40*, 75–81.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping: Effects of training on the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871–888.

Kawakami, K., Dovidio, J. F., & van Kamp, S. (2005). Kicking the habit: Effects of nonstereotypic association training and correction processes on hiring decisions. *Journal of Experimental Social Psychology*, *41*, 68–75.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviours. *Journal of Personality and Social Psychology*, *92*, 957–971.

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.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

LaBar, K. S., Gatenby, C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed trial fMRI study. *Neuron*, *20*, 937–945.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *Journal of Neuroscience*, *15*, 6846–6855.

LaBar, K. S., & Phelps, E. A. (2005). Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behavioural Neuroscience*, *119*(3), 677–686.

Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., & Khan, S. R. (2003). Stereotypes as dominant responses: On the ''social facilitation'' of prejudice in anticipated public contexts. *Journal of Personality and Social Psychology*, *84*, 277–295.

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.

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.

Lehéricy, S., Ducros, M., Van de Moortele, P. F., Francois, C., Thivard, L., Poupon, C., et al. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology*, *55*, 522–529.

Lieberman, M. D., Eisengerger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity to affective stimuli. *Psychological Science*, *18*, 421–428.

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.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1–35.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808–817.

Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews Neuroscience*, *5*, 844–852.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

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

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Cardiovascular reactivity during social interactions with White and Black men. *Personality and Social Psychology Bulletin*, *28*, 939–952.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2008). *Reducing implicit race-biased responses through implementation intentions*. Manuscript submitted for publication.

Millar, M. G., & Tesser, A. (1986). Effects of affective and cognitive focus on the attitude–behaviour relation. *Journal of Personality and Social Psychology*, *51*, 270–276.

Millar, M. G., & Tesser, A. (1989). The effects of affective–cognitive consistency and thought on the attitude-behaviour relation. *Journal of Experimental Social Psychology*, *25*, 189–202.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

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.

Monteith, M. J. (1993). Self-regulation of stereotypical responses: Implications for progress in prejudice reduction. *Journal of Personality and Social Psychology*, *65*, 469–485.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*, 1029–1050.

Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 16, pp. 113–154). Hove, UK: Psychology Press.

Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review*, *2*, 63–82.

Moreno, K. N., & Bodenhausen, G. V. (2001). Intergroup affect and social judgement: Feelings as inadmissible information. *Group Processes and Intergroup Relations*, *4*, 21–29.

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

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*, 167–184.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differently related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Kihlstrom, J. F., D'Esposito, M., et al. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage*, *28*, 797–814.

Ochsner, K., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–249.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421–433.

Park, B., & Judd, C. M. (2005). Rethinking the link between categorisation and prejudice within the social cognition perspective. *Personality and Social Psychology Review*, *9*, 108–130.

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*, 865–870.

Pavlov, I. P. (1927). *Conditioned reflexes*. New York: Dover.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.

Payne, B. K. (2005). Conceptualising control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*, 488–503.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *24*, 27–53.

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

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behaviour. *Neuron*, *48*, 175–187.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, 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.

Pizzagalli, D. A., Sherwood, R., Henriques, J. B., & Davidson, R. J. (2005). Frontal brain asymmetry and reward responsiveness: A source localization study. *Psychological Science*, *16*, 805–813.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832.

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to reduce prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes and Intergroup Relations*, *6*, 187–200.

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

Poldrack, R. A., Selco, S., Field, J., & Cohen, N. J. (1999). The relationship between skill learning and repetition priming: Experimental and computational analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 208–235.

Potanina, P. V., Pfeifer, J. H., Lieberman, M. D., & Amodio, D. M. (2008). *Stereotyping and evaluation in intergroup responses: Dissociable effects of semantic vs. affective memory systems*. Unpublished manuscript.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. K., Pardo, J. V., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, *4*, 8–26.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory*, *1*, 217–229.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F, Wyland, C. L., Trawalter, S., et al. (2004). An fMRI examination 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.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63–131). Amsterdam: Elsevier.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *10*, 284–294.

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z., Nandy, A., & Maddox, K. B. (2007). The effects of skin tone on race-related amygdala activity: An fMRI investigation. *Social Cognitive and Affective Neuroscience*, *2*, 39–44.

Rudman, L. A. (2002). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 79–82.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology*, *81*, 856–868.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalisation: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, *57*, 749–761.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995–1008.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*, 1835–1842.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185–195.

Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes and Intergroup Relations*, *6*, 171–186.

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means. *Psychological Inquiry*, *17*, 173–184.

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

Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science*, *27*, 341–378.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–11562.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

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.

Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Dunn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.

Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 765–779). Cambridge, MA: MIT Press.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, *93*, 13515–13522.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*, 157–175.

van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.

Wagner, A. D., Gabrieli, J. D. E., & Verfaellie, M. (1997). Dissociations between familiarity processes in explicit recognition and implicit perceptual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 305–323.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.

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

Wheeler, M. E., & Fiske, S. T. (2005) Controlling racial prejudice and stereotyping: Social cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*, 56–63.

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

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgements and evaluations. *Psychological Bulletin*, *116*, 117–142.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.