EUROPEAN REVIEW OF SOCIAL PSYCHOLOGY
2011, 21, 237–284

Psychology Press
Taylor & Francis Group

A social neuroscience approach to self and social categorisation: A new look at an old issue

Jay J. Van Bavel,
New York University, NY, New York, USA

William A. Cunningham
The Ohio State University, Columbus, OH, USA

We take a social neuroscience approach to self and social categorisation in which the current self-categorisation(s) is constructed from relatively stable identity representations stored in memory (such as the significance of one's social identity) through iterative and interactive perceptual and evaluative processing. This approach describes these processes across multiple levels of analysis, linking the effects of self-categorisation and social identity on perception and evaluation to brain function. We review several studies showing that self-categorisation with an arbitrary group can override the effects of more visually salient, cross-cutting social categories on social perception and evaluation. The top-down influence of self-categorisation represents a powerful antecedent-focused strategy for suppressing racial bias without many of the limitations of a more response-focused strategy. Finally we discuss the implications of this approach for our understanding of social perception and evaluation and the neural substrates of these processes.

Keywords: Social neuroscience; Social identity; Social categories; Self-categorisation; Social perception; Social cognition; Evaluation; Attitudes; Intergroup relations; Prejudice; Racial bias; Automaticity; New look; Control; Top-down; Salience; Amygdala; Fusiform gyrus; Individuation; Categorisation.

Correspondence should be addressed to Jay J. Van Bavel, New York University, Department of Psychology, 6 Washington Place, room 752, New York, NY, 10003. USA 1003. E-mail: jay.vanbavel@nyu.edu

This manuscript was part of the Jay Van Bavel's PhD Dissertation at the University of Toronto. The authors would like to thank Benjamin Giguere, Jillian Swencionis, Y. Jenny Xiao, Michael Wohl, Miles Hewstone, Wolfgang Stroebe, and four anonymous reviewers for their thoughtful comments on various stages of this manuscript. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada to Jay Van Bavel and the National Science Foundation (BCS-0819250) to William Cunningham.

http://www.psypress.com/ersp DOI: 10.1080/10463283.2010.543314

From a functional perspective, prejudice may be an inevitable aspect of human life (Allport, 1954). Categorising stimuli on the basis of their similarity allows people to manage an otherwise overwhelming amount of incoming information and generalise existing knowledge to new stimuli (Bruner, 1957). People are particularly adept at dividing up the social world into *us* and *them,* and will do so in the absence of factors typically posited to account for intergroup bias, such as stereotypes, prior contact with ingroup or outgroup members and competition over resources (Tajfel, 1970; Tajfel, Billig, Bundy, & Flament, 1971). Indeed, categorising oneself as a member of a group happens in every culture (Brown, 1991). Self-categorisation involves the activation of psychological connections between the self and some class of stimuli (usually other people) at the personal level (i.e., defining oneself as unique from others) or collective level (i.e., defining oneself in terms of similar characteristics with one social group and different characteristics from other social groups) (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987; Turner, Oakes, Haslam, & McGarty, 1994). We propose a social neuroscience framework for understanding how aspects of self-categorisation and social identity shape social perception and evaluation in a top-down fashion.

In the current chapter, we present a social neuroscience framework for social perception and evaluation, outline links between this framework and work on self-categorisation and social identity, and review several experiments showing the utility of this framework for understanding phenomena in social psychology and cognitive neuroscience. We propose that a person's temporary and situation-specific self-categorisation not only influences person perception and evaluation but can even override pervasive racial biases. In the first section, we introduce the iterative reprocessing model (see Cunningham & Zelazo, 2007; Cunningham, Zelazo, Packer, & Van Bavel, 2007) as a framework for understanding how a dynamic self-categorisation may influence social perception and evaluation. In the second section, we review classic research on social identity and social categorisation and introduce our experimental paradigm—a variant of the minimal group paradigm (Tajfel, 1970; Tajfel et al., 1971). In the third section, we review research on automatic social evaluation and studies from our research group demonstrating the influence of self-categorisation and race on social evaluation. In the fourth section, we review research on the neural correlates of person perception and evaluation (the fusiform gyrus and amygdala, respectively) and review data from our research group demonstrating the influence of self-categorisation and race neural activity in these brain regions. Finally we discuss the implications of the presented research in terms of social psychological theory and prejudice reduction.

A SOCIAL NEUROSCIENCE APPROACH TO SELF AND SOCIAL CATEGORISATION

The concept of cognitive control has been central throughout the history of psychology (James, 1896). During much of that time, control has been conceptualised as an intentional ‘‘top-down’’ signal that is often employed to initiate a response or overcome a countervailing, ‘‘bottom-up’’ response tendency. However, top-down controlled processes also modulate the perception of physical objects (Balcetis & Dunning, 2006; Proffitt, 2006). For example, research from the ‘‘new look’’ tradition provided evidence that needs, motives, and expectations can alter perception (Bruner & Goodman, 1947). Thus, top-down control can include both response-focused strategies that suppress or inhibit a response and antecedent-focused strategies that change the perception or construal of a stimulus and/or mental representation (Gross & Thompson, 2007). In the current chapter, we focus on the latter form of top-down influence, showing how shifts in self and social categorisation alter perceptions and automatic evaluations of social stimuli (see also Caruso, Mead, & Balcetis, 2009). This class of automaticity may, in large part, require a person’s consent and intent, similar to *goal-dependent automaticity* (Bargh, 1994).

We examine these processes across multiple levels of analysis, linking the effects of self-categorisation and social identity on perception and evaluation to brain function. This multi-level perspective—variably termed social neuroscience, social cognitive neuroscience, and the social brain sciences—is based on the assumption that complex social phenomena are best understood by combining social and biological theories and methods (Adolphs, 1999; Cacioppo, Berntson, Sheridan, & McClintock, 2000; Ochsner & Lieberman, 2001). This approach involves breaking phenomena like social perception and evaluation into component processes to better understand the operating characteristics of these components and how they work in concert (Cunningham & Van Bavel, 2009; Van Bavel & Cunningham, 2009b). The social neuroscience approach uses current knowledge of the central and peripheral nervous system to test hypotheses about the processes underlying social psychological phenomena and can help develop a functional understanding of the biological systems that underlie social cognition and emotion, which advances science and forms the foundation for future research and intervention (see Ochsner & Lieberman, 2001).[1] Therefore analysing social perception and evaluation across multiple levels of analysis offers the promise of generating more general, process-oriented

[1] Although it is beyond the scope of the current paper, we direct interested readers to a forthcoming issue of *Social Cognition* in which the promise and limitations of social neuroscience are discussed in greater detail.

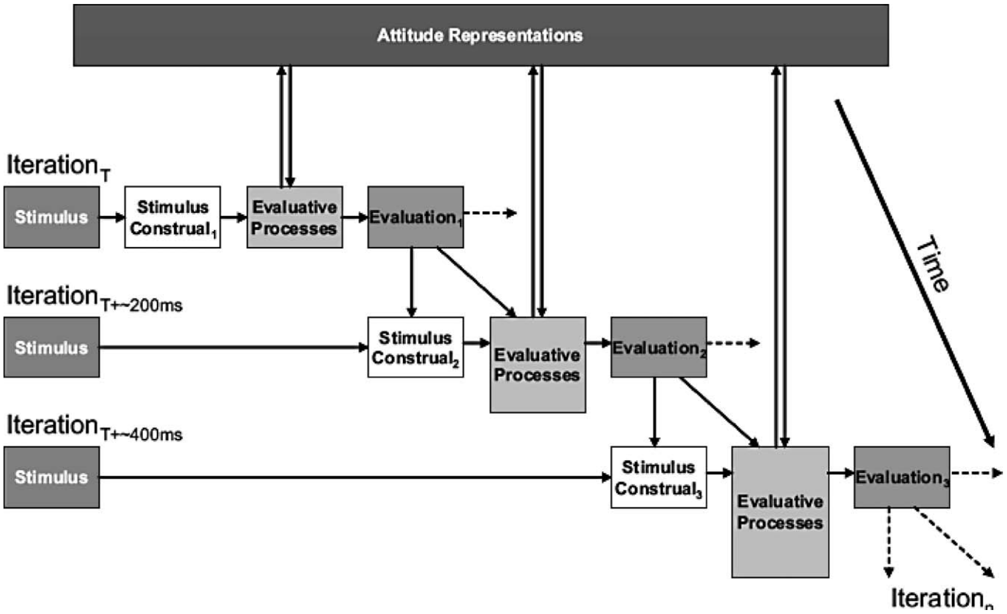

Figure 1. The perceptual and evaluative cycle (reproduced with permission from Cunningham et al., 2007). During each iteration, evaluative processes retrieve attitudinal representations to generate an evaluation relying on a particular construal(s) of the stimulus. This evaluation may influence the next iteration of evaluative processing, direct behaviour, or both. In general, the complexity of evaluative processing (and the resulting evaluation) increases with additional iterations. To view a colour version of this figure, please see the online issue of the Journal.

theories of self and social categorisation and developing novel interventions for pressing social issues like prejudice.

Our approach to these issues is informed by the iterative reprocessing (IR) model of evaluation, a general framework of attitudes and evaluation based on recent advances in social psychology and cognitive neuroscience (Cunningham & Zelazo, 2007; Cunningham et al., 2007). The IR model proposes that current *evaluations* are constructed from relatively stable *attitude* representations stored in memory (this includes evaluative and semantic information, a subset of which are active at any given time) through iterative and interactive evaluative *processing* (see Figure 1).[2] This model provides insight into why automatic evaluations, which are based on

[2] Social psychologists traditionally differentiate aspects of social categorisation (classifying others according to categorical markers), stereotyping (the activation and application of semantic information about others based on their category membership) and prejudice (the evaluation of others based on their social category membership) (Fiske & Neuberg, 1990; Kunda & Sinclair, 1999). Although all three subcomponents often work in concert, their co-occurrence when perceiving others is not necessary. Moreover, while our research focuses on the implications of social categorisation for evaluation, the causal order may be reversed (e.g., Hugenberg & Bodenhausen, 2004).

ostensibly stable underlying associations, appear sensitive to the motivational and social context (see Blair, 2002, for review). Automatic racial bias, for example, is reduced when White participants are placed in a subordinate role relative to a Black partner (Richeson & Ambady, 2003), when they are exposed to admired Black exemplars (Dasgupta & Greenwald, 2001), or when they are in the presence of a Black experimenter (Lowery, Hardin, & Sinclair, 2001). Similarly, categorising complex social stimuli in different ways moderates the activation of underlying automatic attitudes, leading to different evaluations: categorising Black athletes and White politicians according to *race* activates an automatic preference for *White* politicians; however, categorising the same individuals according to *occupation* activates an automatic preference for Black *athletes* (Mitchell, Nosek, & Banaji, 2003). These studies illustrate how stable associations can give rise to contextually sensitive automatic evaluations of social categories.

Building on the distinction between evaluations and attitudes, we integrate core aspects of the IR model with the extant literature on social identity and self-categorisation. Specifically, we propose that the current *self-categorisation* can be constructed from relatively stable representations of a given *identity* (a subset of which are active at any given time) stored in memory. These representations (such as the significance of one's social identity, one's role within the group, etc.) form the current self-categorisation through iterative and interactive perceptual and evaluative *processing*. This allows for the "inherently variable, fluid, and context dependent" nature of self-categorisation (Turner et al., 1994, p. 454) while retaining a set of stable representations of one's personal and social identities. In this way, completing a job interview might make one's professional identity salient without fundamentally altering other, latent identities (such as one's racial or gender identity).

The social context can trigger the activation of aspects of a particular social identity that can, in turn, elicit certain perceptions and evaluations consistent with the contents of that identity. Thus, activating aspects of one's professional identity may lead to activation of occupation-based representations (leading to occupation based categorisations), rather than other contextually irrelevant identities such as race. This would lead to positive evaluations of Black athletes because representations associated with occupation (i.e., athletes) are relatively more likely to be active than the ones associated with race (i.e., Black). Consistent with Self-Categorisation Theory (Turner et al., 1987), we argue that a given "identity" does not reflect the activation of a stable construct stored in memory, but rather the construction of a representation that approximates an identity. The on-line construction is based on a blend of information stored in memory, the incorporation of context and motivation, the computational processes acting on that information, and residual aspects of the perceptual and evaluative systems. Like a snowflake, the same social identity will almost never be activated in an identical fashion twice, even within the same person (O'Reilly & Munakata, 2000). It is also possible to have the representations associated with multiple identities (that lead to different social categorisations) salient at a given time, providing an additive or interactive influence on perception and evaluation (see Crisp & Hewstone, 2007).

A fundamental assumption underlying the IR model is that brain systems are organised hierarchically, such that relatively automatic processes influence and are influenced by relatively more controlled processes in an iterative process. Whereas automatic processes provide relatively coarse perceptual and evaluative information, additional iterations allow for more controlled processes, which can interact with automatic processes and provide more nuanced or contextually appropriate evaluations. This model differs from most dual attitude and/or dual process models in a number of important ways (for a more complete discussion of the similarities and differences please see Cunningham & Zelazo, 2007; Cunningham et al., 2007).

First, the IR model makes no assumption that there are different (implicit versus explicit) attitudinal representations stored in memory (e.g., Rydell & McConnell, 2006; Wilson, Samuel, & Schooler, 2000). According to the IR model, differences in evaluations are largely due to differences in information processing. Second, the IR model does not assume that there are only two qualitatively distinct processes at work in the human mind (e.g., Gawronski & Bodenhausen, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004). Instead, building on recent research on the functional neuroanatomy of the human brain, the IR model assumes there are many, highly interactive neural systems engaged in information processing (see Figure 2). Third, the IR model argues that higher-order controlled processes not only incorporate goals and contexts in the current evaluation, but that prior states of the evaluative system (time $-$ 1) set the stage for automatic construals of the same or subsequent stimuli (at time $=0$). In other words higher-order processes, mediated by top-down control signals from the frontal and parietal networks, can incorporate expectations, goals, bodily states, and contexts into which representations are deemed most relevant in a current context (see Miller & Cohen, 2001), which can then lead to different patterns of self and social categorisation. Thus the preceding context and motivational state of an organism inform ongoing evaluative processes (and vice versa).

Returning to the example above, an individual with a momentarily high need for status might have a higher probability of activating representations associated with his or her occupation as an athlete if he or she is surrounded by adoring fans. The self-categorisation of athlete may be made more or less likely by the nature of the active representations and the current dynamics of

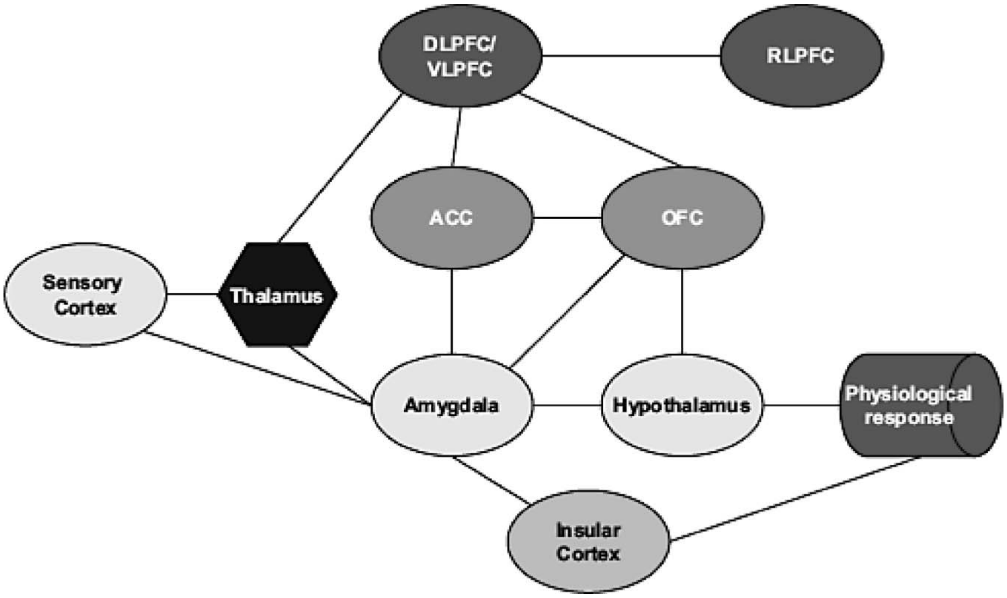

Figure 2. A simplified model of the brain regions underlying evaluation (reproduced from Cunningham et al., 2007). Links between regions discussed in the text are denoted by solid lines (note that not all anatomical links or brain regions are represented). Information about a stimulus may be processed by the thalamus and projected to the amygdala, leading to an initial evaluation that is associated with a tendency to approach or avoid the stimulus. Additional iterations can also include processing by the insula, orbitofrontal cortex (OFC), and anterior cingulate cortex (ACC)—as well as more detailed sensory processing. Visceral changes following evaluation are guided by the hypothalamus and other regions associated with autonomic control. Additional recruitment of the prefrontal cortex, especially regions of the ventrolateral prefrontal cortex (VLPFC), dorsolateral prefrontal cortex (DLPFC), and rostrolateral prefrontal cortex (RLPFC), may subserve goal-oriented reprocessing of stimuli and the regulation of evaluative processing by enhancing or suppressing features of the stimulus or situation.

the system (including a combination of internal and external factors that constrain information processing). As the self-categorisation approaches an equilibrium—the consequence of a settling process where the activation of a self-categorisation is iteratively updated until the change in activations in the network from one cycle to the next is below some threshold (O'Reilly & Munakata, 2000)—the athlete might in turn be more likely to see others in terms of their occupation as well (e.g., opponent, referee, fan, etc.), overriding automatic reactions on the basis of race that some researchers have characterised as inevitable (e.g., Devine, 1989; Fiske, Lin, & Neuberg, 1999; Ito, Willadsen-Jensen, & Correll, 2007) and potentially hard-wired (Olsson, Ebert, Banaji, & Phelps, 2005). In this way the evaluation unfolds over time in a dynamic fashion, reflecting the interactive influence of processes that are both bottom-up and top-down.

It is important to note that our use of the term ‘‘automatic’’ in the present chapter does not imply that evaluations are unintentional, occur outside awareness, are uncontrollable, *and* are efficient in their use of attentional resources (see Bargh, 1984; Johnson & Hasher, 1987). Following the temporally dynamic nature of the IR model, we are primarily interested in the time course of evaluative processes. However, few measures are particularly well suited to examine the time course of perceptual and evaluative processing (with the notable exception of electroencephalography). We therefore rely on previous studies that imply the presence of at least one of the four classic characteristics of automaticity (Bargh, 1994). Many of these previous studies measure evaluations using implicit measures that usually do not require intentionality and are generally efficient, but may be influenced by control (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) and are not necessarily outside conscious awareness (Olson, Fazio, & Hermann, 2007). We provide explicit details about our methods below to clarify the aspects of automaticity that may be implicated in our studies. As we noted above, our discussion of the top-down influence of identity on automatic processing is similar to the concept of *goal-dependent automaticity* (Bargh, 1994), in which the person’s consent and/or intent may be necessary for shaping otherwise automatic responses. In the next section, we review several core concepts from social identity and self-categorisation that have guided the implementation of the IR model in the context of social perception and evaluation.

SOCIAL IDENTITY AND SELF-CATEGORISATION

> Man in his totality is a dynamic complex of ideas, forces, and possibilities. According to the motivations and relations of life and its changes, he makes of himself a differentiated and clearly defined phenomenon. As an economic and political man, as a family member, and as a representative of an occupation he is, as it were, an elaboration constructed *ad hoc*
>
> (Simmel & Wolf, 1950, p. 46)

The value humans place on social categorisation is illustrated by the fact that people form groups and favour ingroup members under rather arbitrary premises. In the classic minimal group paradigm, people who were randomly assigned to a group on the basis of trivial distinctions preferentially allocated money to fellow ingroup compared to outgroup members (Tajfel, 1970; Tajfel et al., 1971). The minimal group paradigm has been replicated numerous times and has reliably produced ingroup favouritism rather than outgroup derogation on a variety of dependent measures (see Brewer, 1979, for a review), including implicit attitude measures (Ashburn-Nardo, Voils, & Monteith, 2001; Otten & Wentura, 1999; see also Perdue, Dovidio, Gurtman, & Tyler, 1990). Whereas *intergroup* bias refers to the tendency to evaluate one's own group (the "ingroup") or its members more favourably than a non-membership group (the "outgroup") or its members in terms of attitude (prejudice), cognition (stereotyping) or behaviour (discrimination; Mackie & Smith, 1998), this group-serving tendency can take the form of favouring the ingroup ("ingroup favouritism") and/or derogating the outgroup ("outgroup derogation"). More importantly, the minimal group paradigm helped lay the foundation for two seminal concepts in intergroup relations: people flexibly categorise themselves in terms of currently salient social groupings (Turner et al., 1987) and social identities are one means by which individuals can fulfil core motives, such as the need for positive distinctiveness (Tajfel, 1982).

The minimal group paradigm highlights the context-dependent nature of self-categorisation and social perception. People not only have many overlapping social identities, but their self-categorisation as a function of these identities—however minimal the distinction between group memberships—can shift from one situation to another (Tajfel, 1982). When perceived group boundaries make one of these social identities salient, people are more likely to perceive themselves and others as interchangeable exemplars of a social category rather than as unique individuals. This self-categorisation, in turn, colours social perceptions and evaluations of the self and others in line with contents of the current self-categorisation (Turner et al., 1987, 1994). This perspective suggests that social perception and evaluation of complex social stimuli are context-dependent and specific to the current salient self-categorisation. Likewise, many studies on person perception suggest that targets are evaluated according to the most significant or salient social category (Macrae, Bodenhausen, & Milne, 1995; Mitchell et al., 2003; Mullen, Migdal, & Hewstone, 2001; Urban & Miller, 1998).

A social identity may shape social cognition and behaviour based on one's knowledge that he or she belongs to that social category. However, social identity reflects not only one's knowledge of his or her membership in a group, but also the value or significance of this group, one's relationship to the group and its members, and the associations one may have with the group and fellow group members (Tajfel, 1972). These characteristics indicate that self-categorisation is but one aspect of the multi-faceted and dynamic nature of social identity. Consistent with the self-categorisation perspective, we propose that social perception and evaluation will reflect the psychologically salient self-categorisation rather than more visually salient social categories, like race. One way to conceptualise this premise is to make a distinction between exogenous and endogenous aspects of perception (see Posner, 1980). In the current context, visually salient categories like race automatically trigger bottom-up, exogenous perceptual processes due to low-level visual features (e.g., physiognomic features). However, a psychologically salient social identity can trigger top-down, endogenous perceptual and evaluative processing, which can attenuate the ostensibly automatic effects of race. Moreover, we propose that the top-down aspects of identity can alter relatively early aspects of perceptual and evaluative processing. This is potentially important, because it introduces the possibility that transient aspects of self-categorisation can override visually salient and socially important social categories—including categories with which people have extensive experience—perhaps before these social categories even begin to influence the perceptual and evaluative system.

We review several studies showing how social identities emerge very rapidly under minimal conditions and override biases in social perception and evaluation that are built on years of social exposure and perceptual expertise. More specifically, we discuss experimental evidence that self-categorisation with a novel mixed-race group organises people's perceptions and evaluations of others in terms of their current self-categorisation rather than race. Although richly developed and important social identities like race may be very important for understanding intergroup attitudes and behaviour, the current review indicates that social perception and evaluation can reflect the current self-categorisation—however minimal—rather than more visually salient identities.

To examine these novel self-categorisations on intergroup processing, we conducted a series of studies in which we assigned participants in these experiments to one of two novel mixed-race teams (Van Bavel & Cunningham, 2009a; Van Bavel, Packer, & Cunningham, 2008). Making race orthogonal to team membership was important for a number of reasons. First, race provided a stringent test of the role of the top-down effects of self-categorisation on social perception and evaluation. As we mentioned above, several researchers have argued that individuals automatically categorise others according to their age, gender, and race, and have considerable difficulty suppressing attention and reactions to those categories, even when those categories are irrelevant to the current task or context (Brewer, 1988; Devine, 1989; Ito & Urland, 2005; Taylor, Fiske, Etcoff, & Ruderman, 1978). Moreover, there is evidence that children categorise and evaluate others according to race as young as 3 years old (Aboud, 1988). Thus, while social categorisation may be flexible, social psychologists frequently report that race is a highly potent and potentially universal social category (Fiske et al., 1999).

The faces on each team were fully counterbalanced and thus visually identical across participants to ensure that any effects of group membership were due to the psychological salience of self-categorisation and not the bottom-up, exogenous properties of different classes of stimuli (see below for the details of each study). Moreover, the top-down effects of self-categorisation would need to override a visually salient, cross-cutting social cue (i.e., race) with well-learned semantic and evaluative associations that have well-documented influences on social perception and evaluation. Second, including race provided a clear test of the role of self-categorisation in a number of racial biases, including perceptual and evaluative racial biases and two key neural correlates of these processes: the fusiform gyrus and amygdala (see below for more details). If these effects are caused by factors other than self-categorisation (e.g., stereotypes, experience, novelty, prejudice, etc.) then race should continue to elicit these biases. However, if our hypotheses are supported, then the top-down influence of the current, salient social categorisation (i.e., membership in a minimal group) should override these biases and it may even suggest that aspects of racial identity are responsible for some of the effects of race (e.g., Sporer, 2001). Finally, including race allowed us to examine the effect of social perception and evaluation in a more complex social context. Although everyone can be categorised and evaluated in a variety of ways, much of the research on social perception and evaluation has focused on simple social contexts with a single salient social category (e.g., comparing reactions to Black and White targets who are otherwise matched in age, gender, etc). Other social categories are usually experimentally controlled or counterbalanced to intentionally reduce the influence of these categories on psychological processing. As shown in Figure 3, we included orthogonal social categories (race and group membership) to empirically examine the effects of multiple social categories on social perception and evaluation (Crisp & Hewstone, 2007). We review a series of experiments on these issues below.

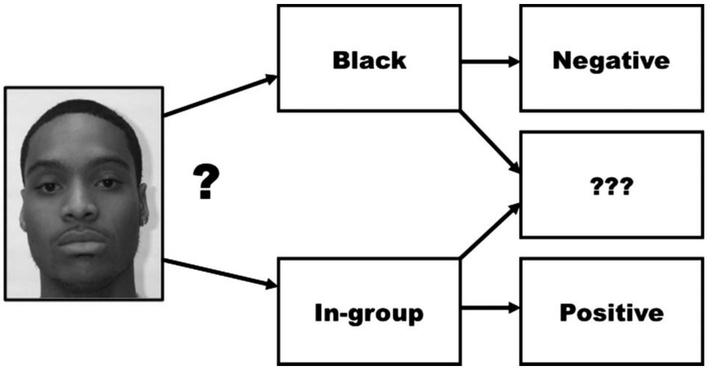

Figure 3. Possible construals and evaluative responses to a Black ingroup target. This target could be categorised according to his race (Black) or group membership (ingroup), which would respectively lead to relatively negative or positive evaluations on an implicit measure. Alternatively, both of these social categories could be integrated to provide a more complex evaluation (e.g., a summation of overall attitudes to the two social categories, an attitude towards a sub-type social category, etc.). (This face in the figure is reproduced with permission from Minear & Park, 2004).

AUTOMATIC SOCIAL EVALUATION: VAN BAVEL AND CUNNINGHAM (2009)

Dating to the early twentieth century, scholars began to notice a growing tension between racial bias and the widely held belief in equality. Economist Gunnar Myrdal (1944) described this ''ever-raging conflict'' as the American Dilemma. Different from bigots or old-fashioned racists who feel wholly justified in their prejudices, conflicted forms of prejudice became an increasingly common experience among White Americans who pay for their prejudices with guilt or compunction (Allport, 1954). In fact, racial prejudice and discrimination are evident (i) when they are measured through unobtrusive or subtle means (Crosby, Bromley, & Saxe, 1980), (ii) when the social norms against the expression of prejudice are ambiguous (Gaertner & Dovidio, 1977), or (iii) when there is competition over limited resources (Levine & Campbell, 1972). These biases are even expressed (usually in a more subtle fashion) by people who explicitly endorse egalitarian values, including those who genuinely believe they are non-prejudiced (Gaertner & Dovidio, 1986; Pettigrew & Meertens, 1995).

The persistence of racial bias in the presence of egalitarian values has led researchers to distinguish between automatic and controlled processes in social prejudice. There is now extensive evidence that social categorisation and evaluation occur automatically (Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). These initial insights were bolstered by the development of several implicit measures (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Payne, Cheng, Govorun, & Stewart, 2005). These measures provided experimental evidence that the majority of White Americans appear to have at least some automatic racial bias against Blacks (Nosek, Banaji, & Greenwald, 2002), suggesting that these biases are remarkably pervasive. Presumably, the associations that people have with race are often so well learned that they are automatically activated upon encountering members of these groups (Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998). Moreover, people with stronger automatic racial bias display more subtle, non-verbal discrimination in interracial interactions (Dovidio, Kawakami, & Gaertner, 2002; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; McConnell & Leibold, 2001).

Various dual process models have described how automatic perceptions and evaluations that occur rapidly and without intent are often in conflict with slower evaluations that reflect current goals and motivations (see Chaiken & Trope, 1999; Greenwald & Banaji, 1995; Wilson et al., 2000). In the context of race, Devine (1989) proposed that the activation of stereotypes occurs without intention, effort, or conscious control, and regardless of personal prejudice towards a group, making it a virtually unavoidable aspect of intergroup perception. Accordingly, dissociations between automatic and controlled processes are especially likely when people have the motivation and opportunity to express an evaluation at odds with their initial reaction (Fazio & Towles-Schwen, 1999).

One of the reasons race may serve as such a powerful trigger for social prejudice and discrimination is because it provides a cue to group membership in the United States. In social environments in which there is less than complete racial integration, race may provide a visually salient cue to group membership (Cosmides, Tooby, & Kurzban, 2003; Sidanius & Pratto, 1999). However, any individual can be categorised according to multiple dimensions, including age, gender, race, occupation, and nationality (Crisp & Hewstone, 2007; Deschamps & Doise, 1978; Mullen et al., 2001; Urban & Miller, 1998) and the psychological salience of any given social category can shift relatively quickly. Research on the common ingroup identity model, for example, has shown that categorising two separate groups (us and them) into an inclusive group (we) reduces self-reported intergroup bias (Gaertner, Rust, Dovidio, Bachman, & Anastasio, 1996). Alternatively, making multiple, cross-cutting social categories salient can lead people to perceive a shared social identity with outgroup members, reducing self-reported intergroup bias (Crisp & Hewstone, 2007). When race is unrelated to another dimension of group membership, this other dimension may drive social perception and evaluation (Cosmides et al., 2003; Kurzban, Tooby, & Cosmides, 2001; Sidanius & Pratto, 1999).

We therefore conducted a pair of studies in which participants were assigned to mixed-race groups and then completed both automatic and controlled measures of their attitudes towards ingroup and outgroup members (Van Bavel & Cunningham, 2009a). Participants arrived at the lab and posed for a digital photograph. They were then informed that they were in a study exploring how people learn about groups, and randomly assigned to one of two groups (the Lions or Tigers) or a control condition in which they learned about the two groups without being assigned to one of them. Participants then completed two brief learning tasks ($\sim$15 minutes) in which they memorised the group membership of 24 faces. Participants saw two mixed-race groups in which six Black and six White males were in each group. In the first learning task, participants spent 3 minutes memorising the group membership of all 24 faces simultaneously: 12 members of Lions and 12 members of the Tigers. In the second learning task, participants were presented with each of the 24 faces one at a time and indicated whether each face was a member of the Lions or Tigers. We randomly assigned the faces to group and fully counterbalanced them so that nothing in the appearance of the individuals allowed participants to visually sort them into groups. This design logically guaranteed that there were no visual differences between group members across participants.

Finally, participants completed automatic and controlled evaluation measures (counterbalanced) of the faces without explicit category labels.

Previous research has shown the salience of difference social categories moderates the automatic activation of underlying attitudes. Thus, when participants categorise Black athletes and White politicians according to *race*, this activates an automatic preference for *White* politicians, whereas categorising the same targets according to *occupation* activates an automatic preference for Black *athletes* (Mitchell et al., 2003). However, the automatic evaluations of targets in these studies may have been driven by attitudes towards the category *labels* rather than the spontaneous construal of the targets, because some implicit attitude measures, including the Implicit Association Test (Greenwald et al., 1998), evoke evaluations consistent with the category labels (Olson & Fazio, 2003). Moreover, the salience of these category labels would make it difficult in our studies to determine whether targets are spontaneously evaluated according to pre-existing racial bias (White > Black), current salient self-categorisation (ingroup > outgroup), the sum of these categories, or an interaction between race and self-categorisation (see Crisp & Hewstone, 2007). As seen in Figure 3, it was possible that a Black ingroup member would be categorised as Black, leading to a relatively negative evaluation, an ingroup member, leading to a relatively positive evaluation, or some combination of these categories. To allow for these possibilities, we measured automatic evaluations of the faces using a computerised response-window priming task (Cunningham, Preacher, & Banaji, 2001; Draine & Greenwald, 1998).

During this task participants were instructed to rapidly categorise a word on each trial as "good/liked" or "bad/disliked" (see Olson & Fazio, 2004). On each trial a face appeared on the centre of the computer monitor for 150 ms (followed by a blank screen for 50 ms) before a positive (e.g., love) or negative (e.g., hatred) target word, which appeared for 525 ms (see Figure 4). Participants were instructed to press "1" when a good word appeared and "2" when a bad word appeared. Since longer response times allow for more controlled processing (Neely, 1977), we only analysed responses that occurred within 600 ms. Importantly, participants were instructed to focus on the word-categorisation task and completely ignore the faces. The dependent measure was the proportion of trials in which each participant correctly categorised the word as good or bad. We assumed that faces with positive associations (e.g., ingroup faces) would increase accuracy to positive words and decrease accuracy to negative words. In contrast, faces with negative associations (e.g., Black faces) would decrease accuracy to positive words and increase accuracy to negative words. It is worth noting that Black faces (or faces from most social categories) are associated with both positive and negative information. However, the net association is generally negative among White Americans (e.g., Nosek et al., 2002). More

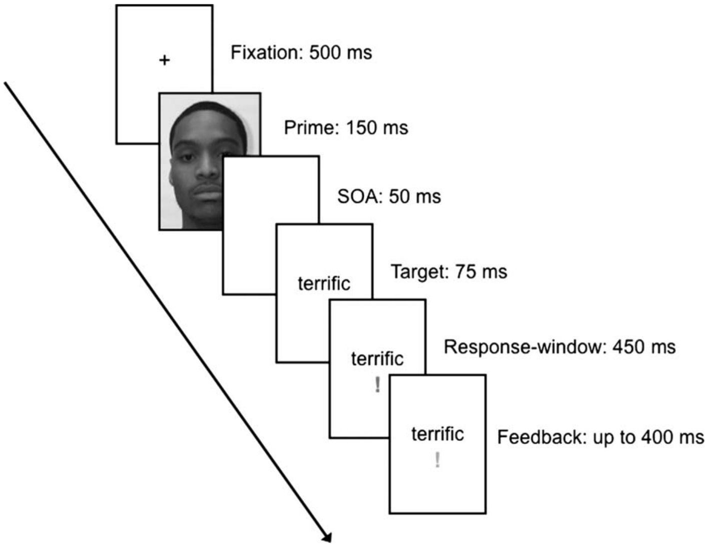

Figure 4. A sample trial in the response-window priming paradigm used in Van Bavel and Cunningham (2009b). So that we could measure automatic evaluations of the faces, participants completed a response-window priming task on a personal computer (Draine & Greenwald, 1998). During this task participants were instructed to rapidly categorise each word as "good/liked" or "bad/disliked" (Olson & Fazio, 2004). Participants were instructed to press 1 when a good word appeared and 2 when a bad word appeared. Following 24 practice trials, participants completed three critical blocks with 96 trials each. On each trial in these critical blocks a face from the learning task appeared for 150 ms (followed by a blank screen for 50 ms) before a positive (e.g., love) or negative (e.g., hatred) target word, which appeared for a 525-ms response window. Participants were instructed to ignore the faces. The dependent measure was the proportion of trials in which each participant correctly categorised the word as good or bad. To better estimate automatic evaluative processing, all responses that occurred after 600 ms were coded as incorrect because longer response times allow for more controlled processing (Neely, 1977) and we were interested in automatic evaluation. We assumed that faces with positive associations would increase accuracy to positive words and decrease accuracy to negative words.

importantly, we predicted that automatic evaluation of targets would reflect the current self-categorisation (i.e., members of a mixed-race team) such that participants would show an automatic preference for minimal ingroup members even when race was an orthogonal, visually salient social category.[3] We assumed that assigning people to a mixed-race group would

[3] It is important to note that the current experiments employed a modified version of the minimal group paradigm (Tajfel et al., 1971): to enhance self-categorisation participants were told that the Lions and Tigers were in competition and saw their own face appear during the learning task. In addition, participants actually had to learn ingroup and outgroup faces prior to completing the dependent measures. Although this variant of the minimal group paradigm departed from the classic version, we have replicated these results in follow-up studies in which participants did not see their own face and in which there was no reference to competition. We therefore feel confident in loosely describing the groups in these studies as minimal groups.

TABLE 1
Mean accuracy as a function of word valence (positive, negative), race (black, white), and group membership (ingroup, outgroup) (Van Bavel & Cunningham, 2009a)

	Accuracy during the response-window priming task			
	White		*Black*	
Task	*Positive*	*Negative*	*Positive*	*Negative*
Experiment 1: Control condition				
All faces	.82 (.02)	.78 (.02)	.83 (.02)	.83 (.02)
Experiment 1: Experimental condition				
Ingroup	.84 (.01)	.81 (.01)	.85 (.01)	.81 (.01)
Outgroup	.85 (.01)	.83 (.01)	.83 (.01)	.84 (.01)
Experiment 2: Experimental condition				
Ingroup	.78 (.02)	.78 (.02)	.78 (.02)	.76 (.02)
Outgroup	.77 (.02)	.75 (.02)	.78 (.02)	.78 (.02)
Unaffiliated	.78 (.02)	.77 (.02)	.75 (.02)	.78 (.02)

Accuracy = the proportion of trials with correct response during the response-window priming task. Excludes all trials where the reaction time < 300 ms. The estimated least squared means from multi-level models are presented with standard errors in parentheses. Differences between conditions may be distorted due to rounding.

alter the salience of participants' current self-categorisation and exert a top-down influence on evaluation, even on automatic measures.

As predicted, participants assigned to a mixed-race group had positive automatic and controlled evaluations of Black *and* White ingroup members, and these evaluative preferences were driven by ingroup favouritism and not outgroup derogation (see Table 1). In other words, group membership increased relative positivity (positive – negative) towards Black ingroup members relative to Black outgroup members, eliminating the standard pattern of automatic racial bias (Cunningham et al., 2001; Fazio et al., 1995). In contrast, uncategorised participants who merely saw two mixed-race groups, without being assigned to one of them, showed the standard pattern of automatic racial bias (more positive evaluations of White compared to Black faces), highlighting the power of self-categorisation and social identification to shape automatic evaluation. These experiments provide evidence that automatic evaluation is highly sensitive to the top-down influence of self-categorisation. Participants' preferences reflected their current salient self-categorisation even when there were no visual differences between the ingroup and outgroup (since the members of each group were fully counterbalanced across participants), the groups have no

However, it does remain an open question whether mere categorisation is sufficient to override automatic racial bias (see Van Bavel & Cunningham, 2009a, for a discussion).

history of contact or conflict, and there is an orthogonal, visually salient social category cue (i.e., race) with strong evaluative connotations. These results also show that mere categorisation with a relatively unimportant group is sufficient to override automatic evaluations of ingroup members according to race.

We used implicit and explicit measures of evaluation without explicit category labels in this experiment. Several previous studies on automatic evaluations have focused on relatively simple social categories or examined the automatic evaluations of complex social targets with measures that use explicit category labels. As we noted above, participants who were forced to categorise Black athletes and White politicians as Black versus White on the Implicit Association Test had more pro-White/anti-Black racial bias than participants who were forced to categorise the same stimuli as athletes versus politicians (e.g., Mitchell et al., 2003). In both of our own experiments we found that the automatic evaluations of ingroup and outgroup faces were complex, involving an interaction between the minimal group membership and race. Participants had a preference for ingroup members, regardless of race, but continued to show automatic racial bias towards outgroup members. This pattern of automatic evaluations suggests that people spontaneously and rapidly evaluate others according to a blend of psychologically salient self-categorisations and pre-existing associations towards visually salient social categories.

Taken together, these results raise the possibility that participants were either generating a superordinate social identity that included all Black and White targets (except for outgroup members) (Gaertner et al., 1996) or that the crossed-categorisation of novel group membership and race led to this reduction in racial bias (Crisp & Hewstone, 2007). If participants were generating a superordinate identity they should show little or no racial bias towards any Black individual who is not an outgroup member. To examine this issue directly, participants in a follow-up study were presented with Black and White faces unaffiliated with the ingroup or the outgroup during the evaluation tasks (Van Bavel & Cunningham, 2009a). This experiment was similar to the previous experiment with the primary difference that participants also evaluated Black and White faces that were unaffiliated with either mixed-race group. Specifically, participants memorised 8 members of the Tigers and 8 members of the Lions (instead of 12 members of each team) during the learning phase. Participants then evaluated these 16 faces along with 4 Black and 4 White faces that were unaffiliated with the Lions or Tigers. Participants saw these unaffiliated faces for the first time during the evaluation tasks. Contrasting the evaluations of unaffiliated faces with ingroup and outgroup members allowed us to determine whether the preference for Black ingroup compared to Black outgroup members was driven by ingroup favouritism or outgroup derogation.

Interestingly, participants in this experiment revealed automatic racial bias towards Black faces that were unaffiliated with the ingroup or outgroup (see Table 1). In other words, the relatively positive (positive – negative) evaluation of Black faces was limited to ingroup members and did not generalise to unaffiliated faces, suggesting that participants were not generating a superordinate ingroup identity, but one specific to their minimal ingroup (Lions or Tigers). It seems more likely a hierarchical identity structure, in which race is nested within group membership, may have accounted for the reduction in racial bias (see Van Bavel & Cunningham, 2009a). We also found that Black ingroup faces were evaluated more positively than unaffiliated Black faces which were not evaluated differently than Black outgroup faces. This pattern of results provided evidence that our results were best characterised as ingroup favouritism and not outgroup derogation. In the next section, we describe the role of specific psychological processes associated with brain regions implicated in social perception and evaluation (the fusiform gyrus and amygdala, respectively) using experimental paradigms and convergent behavioural evidence.

NEURAL SUBSTRATES OF SOCIAL PERCEPTION AND EVALUATION: VAN BAVEL ET AL. (2008)

Nearly half a century of research has explored how social categories alter social perception. One of the most robust and widely replicated phenomena in social perception is the fact that people appear to be better at remembering people from their own race than from other races (Malpass & Kravitz, 1969)—an effect that has been variably termed the cross-race effect, same-race bias, or own-race bias (ORB). This simple psychological phenomenon has caused countless individuals to exclaim that members of another race or ethnicity ‘‘all look the same to me’’, providing fodder for cartoonists, comedians, and satirical websites (e.g., alllooksame.com). Although the ORB may appear to be a relatively innocuous error, it can lead an eyewitness in a criminal case to misidentify a suspect from another race, leading to the conviction of an innocent person (Brigham & Ready, 2005).

For the past four decades, perceptual expertise has been widely accepted as the primary psychological mechanism for ORB. According to this account, people become expert at identifying individuals within their own race by virtue of their exposure to own-race individuals, including family, friends, and acquaintances, relative to members of another race, which produces a specific expertise for encoding and/or recalling own-race faces. Over the course of a lifetime of interactions with own-race members, people become expert at making within-race distinctions on the basis of physiognomic features and making between-race distinctions by contrasting features that distinguish own-race from other-race individuals (Malpass & Kravitz, 1969). This combination of experience in making both within- and between-race distinctions tunes the perceptual system to distinguish more among exemplars within own-race faces than within other-race faces. Consistent with the expertise model, people generally report greater experience with own-race faces (e.g., Malpass & Kravitz, 1969), and several studies have reported a correlation between own-race expertise or contact and ORB (Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). Despite the intuitive appeal of expertise models and evidence that expertise enhances recognition memory in other domains (McClelland & Chappell, 1998), the empirical support for the expertise model of ORB is actually mixed. Some studies have shown that lifelong experience with own-race faces is associated with the ORB (Sangrigoli et al., 2005) while others have found no relationship between interracial contact (a proxy for expertise) and the ORB (Ng & Lindsay, 1994). Indeed, interracial contact accounts for only 2% of the total variance of the ORB (Meissner & Brigham, 2001), which is considered a small effect size (Cohen, 1988). The lack of strong empirical support for the expertise model suggests that other models might provide a more comprehensive account of the ORB.

In the past decade, social categorisation approaches have challenged the perceptual expertise model of ORB (Levin, 2000; Sporer, 2001). In a comprehensive review of the ORB literature, Sporer (2001) proposed the ingroup–outgroup model to account for the ORB (see Figure 5). According to this model, categorising others as ingroup or outgroup members may alter the depth or type of processing that they receive, such that own-race faces are processed as individuals by default and other-race faces are processed as interchangeable representatives of a social category, leading to superior recognition memory for own-race faces (Bernstein, Young, & Hugenberg, 2007; Levin, 1996, 2000; see also the outgroup homogeneity effect: Ostrom & Sedikides, 1992; Sporer, 2001). This approach is similar to models of person perception wherein people tend to think categorically about outgroup members, relying on stereotypes about their social category membership (e.g., age, gender, race) to inform evaluations and judgements, whereas they individuate ingroup members due to their personal/motivational relevance, and therefore use individual characteristics to inform their evaluations and judgements (Brewer, 1988; Fiske & Neuberg, 1990).

A functional magnetic resonance imaging (fMRI) study (Golby, Gabrieli, Chiao, & Eberhardt, 2001) examined the relationship between the ORB and the activation in the Fusiform Face Area (FFA), a sub-region of the fusiform gyrus located on the ventral surfaces of the temporal lobe (see Figure 6). The FFA is in a slightly different place for every person, but tends

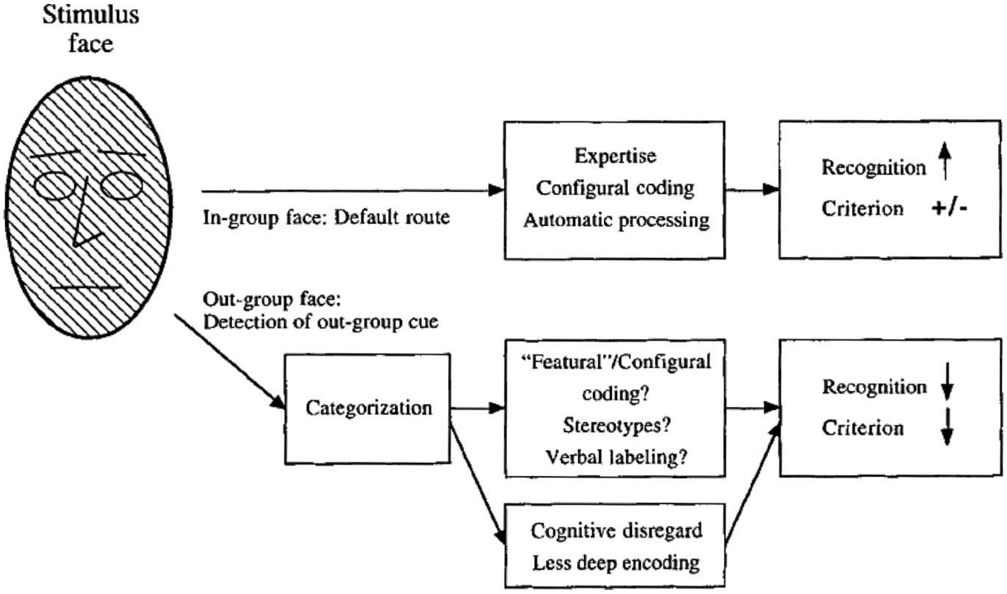

Figure 5. Image of Sporer's ingroup/outgroup model of face processing (reproduced from Sporer, 2001). According to this model, categorising others as ingroup or outgroup members alters the depth or type of processing that they receive, such that ingroup faces are processed as individuals by default and outgroup faces are processed categorically, leading to superior recognition memory for ingroup or own-race faces.

to be lateralised in the right hemisphere in most people. Several dozen studies have shown that the FFA responds preferentially to faces relative to other objects (Kanwisher, McDermott, & Chun, 1997; Sergent, Ohta, & MacDonald, 1992) and lesions to this region lead to prosopagnosia, a deficit in face recognition that spares the ability to recognise non-face objects (Benton & Van Allen, 1968; De Renzi & Spinnler, 1966; Ellinwood, 1969). These studies led Kanwisher and colleagues to argue that the FFA reflects a specialised mechanism for detecting the presence and identity of faces. There nevertheless remains considerable debate about the nature of information processing in this region. The most prominent challenge to the face-specific hypothesis has come from several neuroimaging studies, which have provided evidence that FFA activity increases with visual expertise (see Palmeri & Gauthier, 2004, for a review). For example, car and bird experts have heightened FFA activity while viewing cars and birds, respectively, than familiar objects, such as furniture (Gauthier, Skudlarski, Gore, & Anderson, 2000a). Further, Gauthier and colleagues created expertise to novel stimuli called greebles, and found greater FFA activity during the passive viewing of greebles among trained greeble experts (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). These studies led Gauthier and colleagues (1999) to suggest that the FFA is better termed the *flexible fusiform area*, because processing in the region is not limited to pre-

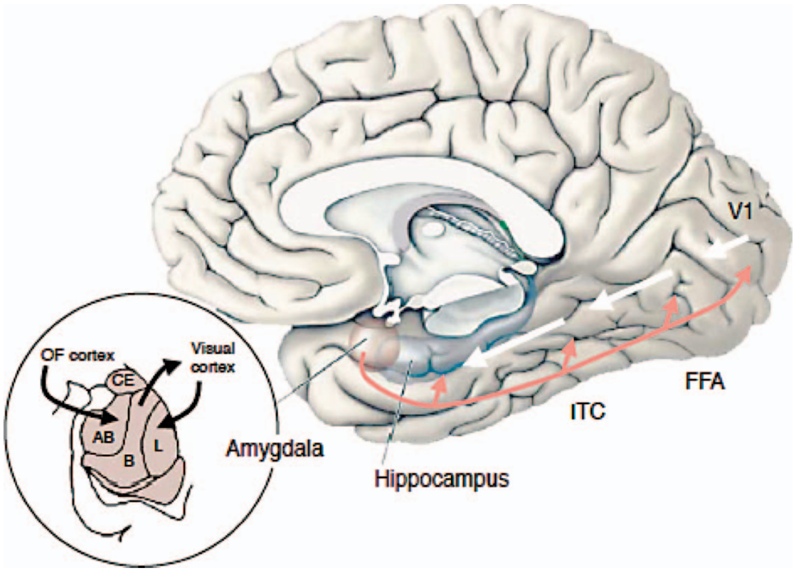

Figure 6. A visual representation of the anatomical location of and relationship between the amygdala and fusiform gyrus (adapted from Vuilleumier, 2005). The image displays feedback connections between amygdala and visual cortex, including the fusiform gyrus face area (FFA). As shown in the figure, the amygdala not only receives inputs from ventral visual cortical pathways (in its lateral nucleus; L), but also sends connections (from its basal nucleus; B) back to virtually all regions along the ventral visual processing stream, including the inferior temporal cortex (ITC) and primary visual cortex (area V1). (AB, accessory basal nuclei; CE, central nuclei; OF, orbitofrontal).

determined content, such as faces. However, the flexibility of this region appears to be constrained by requirements for visual expertise. For example, greeble expertise requires over 3000 trials of intensive training over several days or weeks (Gauthier et al., 1999; Gauthier, Williams, Tarr, & Tanaka, 1998). These studies suggest that extensive visual experience with faces or other stimulus categories gradually tunes neurons in the FFA to encode stimuli at the sub-ordinate/individual level (Tarr & Gauthier, 2000).[4]

Building on this research, Golby and colleagues (Golby et al., 2001) presented Black and White participants with pictures of Black and White faces as well as objects (radios) during neuroimaging. Brain activity to the faces was first contrasted with objects to identify the location of the FFA in each participant (Kanwisher et al., 1997). As the authors predicted, activity in FFA was greater to own-race than other-race faces for both Black and White participants (see also Lieberman, Hariri, Jarcho, Eisenberger, &

[4] There is some evidence that perceiving others as specific individuals does not lead to enhanced FFA activity (Kriegeskorte, Formisano, Sorger, & Goebel, 2007). Nevertheless we use the term individuation to reflect the in-depth structural analysis of faces that is well established within the FFA literature (Kanwisher & Yovel, 2006).

Bookheimer, 2005). Moreover, on a subsequent memory test, the degree of same-race bias (i.e., superior memory for same-race over other-race faces) was predicted by fusiform gyrus activation to racial ingroup members. These experiments suggest that extensive visual experience with faces or other stimulus categories, including one's race, may gradually tune neurons in the FFA to encode stimuli at the subordinate/individual level: that is, to make fine-grained discriminations between exemplars within a stimulus category (Tarr & Gauthier, 2000).

However, expertise may not be necessary to distinguish between exemplars within a category (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997). For example, participants who completed a task in which they matched non-face stimuli with superordinate categorical (e.g., bird) versus subordinate-level (e.g., pelican) descriptors had greater activity in the ventral visual pathway, including the fusiform gyrus, during subordinate level judgements (Gauthier et al., 1997, 2000b). Thus it seems that activity in the fusiform gyrus may not be fully contingent on expertise with specific categories but rather that it might be sensitive to top-down motivational factors. We therefore predicted that participants assigned to a novel minimal group would encode ingroup members at a subordinate level and outgroup members at a more superordinate level, and that these differences in encoding would be reflected in differences in fusiform activity (ingroup > outgroup), despite participants' limited exposure to members of both categories.

To examine the role of the fusiform gyrus in social perception, we assigned participants to a mixed-race minimal group (Van Bavel et al., 2008). Assigning participants to a social category through an experimental procedure *rather* than using an existing intergroup distinction (e.g., race) allowed us to examine fusiform activity in the absence of differential exposure to ingroup/outgroup members or visual cues that signify group membership. Similar to the studies discussed above, we informed our White participants that they were in a study exploring learning about groups and that they had been assigned to the Leopards or Tigers for the duration of the study. We explained that it was important for them to learn the members of their group and the other group before moving to other phases of the study. Participants then completed two brief learning tasks. During the first learning task, participants spent 3 minutes memorising the team membership of 24 faces presented simultaneously: 12 members of the Leopards and 12 members of the Tigers. Race was orthogonal to team membership; there were six Black and six White males on each team. Faces were randomly assigned to team and fully counterbalanced so that participants were equally likely to see each face as an ingroup or outgroup member, and nothing in the appearance of the individuals allowed participants to visually sort them into teams.

During the second learning task, participants saw and categorised each face according to whether the face was affiliated with the Leopards or Tigers. It is also important to note that we took a digital photograph of each participant when they arrived at the scanning centre, and they saw and categorised their own face three times during the second learning task. During the first block of trials participants were reminded with a label on the screen whether each face was a Leopard or Tiger. During the second block of trials the label was removed so that participants needed to rely only on their memory. Following each trial, feedback indicated if the response was correct. After the learning phase participants were presented with the same 24 Black and White faces they had seen during the learning phase. Specifically, they saw and responded to six White and six Black ingroup faces and six White and six Black outgroup faces during neuroimaging. We utilised a rapid event-related design in which participants completed 288 trials in which they categorised one of the 24 faces according to team membership (Leopard or Tiger) or skin colour (Black or White). Each face appeared for 2 seconds, during which time participants responded with a button-box in their right hand. It is worth noting that participants did not see or respond to their own face during the neuroimaging task, and they never saw or interacted with any member from either team—their "experience" with ingroup and outgroup members was limited to brief exposure to facial photographs.

We reasoned that if the fusiform gyrus is merely processing expert stimuli, the White participants would show greater activity to White relative to Black faces. However, if the fusiform is involved in individuation, participants would show greater activity to ingroup relative to outgroup faces, even when the intergroup distinction is arbitrary and exposure to ingroup and outgroup faces is equivalent, brief, and very recent. Moreover, this effect should occur regardless of the race of ingroup members. As predicted, there was greater mean activity within the bilateral fusiform gyrus for ingroup than outgroup faces (see Figure 7). These results provide convergent evidence that the fusiform gyrus is sensitive to shifts in social context, responding selectively to face stimuli that are imbued with psychological significance by virtue of their group membership, encoding the more motivationally relevant ingroup faces at the subordinate level. Moreover, these effects were not moderated by race (nor was there a main effect of race; see also Hehman, Maniab, & Gaertner, 2010; Kinzler, Shutts, DeJesus, & Spelke, 2009; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008). Indeed, our group assignment manipulation ensured that no perceptual cues allowed participants to visually sort the faces into teams. Only the experimental manipulation of group membership could account for the difference in fusiform activity between ingroup and outgroup faces.

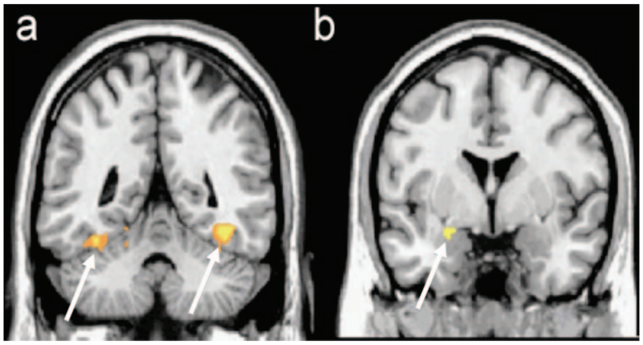

Figure 7. Images of brain areas in which activity was greater for ingroup than for outgroup faces (reproduced from Van Bavel et al., 2008). Areas showing this effect included (a) the fusiform gyri (coronal view; y = –48), (b) the amygdala (coronal view; y = 0). To view a colour version of this figure, please see the online issue of the Journal.

Several papers have argued that the fusiform gyrus may play a central role in individuating faces (George et al., 1999; Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher & Yovel, 2006; Winston, Henson, Fine-Goulden, & Dolan, 2004), and perhaps a general role in subordinate-level processing (Palmeri & Gauthier, 2004). Our research provides evidence that the motivational relevance of categories, like group membership, can affect fusiform activity in a flexible and dynamic fashion in the absence of long-term experience with the category or explicit task instructions. By virtue of their motivational significance in a variety of contexts (e.g., economic, psychological, and evolutionary), ingroup members often warrant greater/deeper processing than outgroup members (Brewer, 1979, 1999). We believe our study suggests that the fusiform may play a key role in processing ingroup members in greater depth than outgroup members—placing ingroup biases in perception firmly within the realm of motivated social perception (Balcetis & Dunning, 2006).

Indeed, we have recently replicated this pattern of ingroup bias in the FFA (a functionally localised sub-region of the fusiform that is sensitive to faces) and shown that relatively greater activity in this region mediates the effects of group membership on recognition memory—a behavioural index of individuation (Van Bavel, Packer, & Cunningham, 2010). Specifically, we found a positive correlation between the brain activation differences in the FFA for ingroup and outgroup faces and recognition memory differences for ingroup and outgroup faces. Thus the minimal ingroup/outgroup distinction not only organises brain activity in a fashion that appears to override the effects of race, but this difference may reflect more than a mere shift in categorisation focus (see also Bernstein et al., 2007; Hehman et al., 2010; Kinzler et al., 2009). Our results imply that ingroup members are more likely to be processed as individuals or exemplars in a non-categorical fashion than outgroup members, consistent with social cognitive models of person perception (Brewer, 1988; Fiske & Neuberg, 1990; Sporer, 2001). In the next section we examine the amygdala, a region in the extended face network (Bruce & Humphreys, 1994; Haxby, Hoffman, & Gobbini, 2000) that plays an important role in social evaluation (see Macrae & Quadflieg, 2010).

The central focus of social neuroscience research on social prejudice has been the affective response people have to race. Although the neural networks involved in an affective evaluative response are widely distributed (Cunningham et al., 2007), initial research focused on a small structure in the temporal lobe called the amygdala (see Figure 6). The amygdala has been implicated in a host of social and affective processes (for a review see Phelps, 2006), including fear conditioning (LeDoux, 2000), processing negative stimuli (Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003; Hariri, Tessitore, Mattay, Fera, & Weinberger, 2002), and perceiving emotional faces (Whalen et al., 1998). More strikingly, the amygdala is activated during rapid subliminal presentations (Whalen et al., 1998). In these studies the parallels between amygdala activity and dual process models of racial bias suggest the amygdala may mediate automatic racial bias.

Initial studies on the neural substrates of social prejudice examined amygdala activity to faces from different racial groups. Viewing images of other-race faces members activates the amygdala more than does viewing own-race faces (Hart et al., 2000). Macrae and Quadflieg (2010) have argued that the amygdala seems to respond to a target's race when participants are making socially meaningful age or sex judgements (Hart et al., 2000; Ronquillo et al., 2007; Wheeler & Fiske, 2005), but not when targets are processed at a superficial visual level, as during simple perceptual tasks (Cunningham et al., 2004; Phelps et al., 2000; Wheeler & Fiske, 2005). However, Lieberman and colleagues (2005) have shown the exact opposite pattern of results: Black and White participants both show greater amygdala activity to Black compared to White faces when these faces are matched on perceptual features, but this pattern of racial bias is reduced when participants engage in explicit racial judgements. Moreover, Cunningham and colleagues (2004) find greater amygdala activity to Black compared to White faces in a simple localisation task when the faces are presented subliminally. Thus, while the amygdala is clearly sensitive to race, the distinction between deep and superficial judgements does not fully characterise the relationship between race and amygdala activity.

Other researchers have examined whether automatic and controlled processes might provide a more useful framework for understanding the relationship between race and amygdala activity. Several studies have now shown that individual differences in amygdala activity for Black compared to White faces correlate with implicit measures of racial bias (Cunningham et al., 2004; Phelps et al., 2000), including the Implicit Association Test (Greenwald et al., 1998) and the startle eye-blink (a physiological measure) (Amodio, Harmon-Jones, & Devine, 2003). These correlations with racial bias, coupled with studies demonstrating a link between the amygdala and fear conditioning (LeDoux, 1996), have led researchers to interpret differences in amygdala activation in intergroup contexts as evidence of negativity (including disgust and fear) towards stigmatised groups (e.g., Harris & Fiske, 2006; Krendl, Macrae, Kelley, & Heatherton, 2006; Lieberman et al., 2005). However, differences in amygdala activity to different race faces are generally uncorrelated with more explicit measures of prejudice (Phelps et al., 2000), like the Modern Racism Scale (McConahay, 1986). The dissociation between implicit and explicit measures of racial bias and amygdala activity is consistent with numerous dual process models of social prejudice.

According to most contemporary models of evaluation, people can control their automatic responses, and generate different or more nuanced evaluations and judgements in the service of their goals and values (Cunningham & Zelazo, 2007; Fazio, 1990; Greenwald & Banaji, 1995). The study of prejudice regulation in social psychology and in social neuroscience has tended to focus on the inhibition or suppression of automatic evaluations deemed inappropriate or inaccurate (Devine, 1989; Petty & Wegener, 1993). Specifically, when people have the motivation and opportunity to engage in more controlled processing, the influence of automatically activated stereotypes and prejudice is normally reduced (Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995). In other words, the automatic activation of prejudiced representations and biased processes leads to discriminatory behaviour unless controlled processes driven by goals and motivations attenuate these biases.

Research in cognitive neuroscience has identified at least two separate, but related, neural systems involved in controlled processing: a conflict-detection and a regulatory-control system (Botvinick, Braver, Barch, Carter, & Cohen, 2001; MacDonald, Cohen, Stenger, & Carter, 2000). The conflict-detection system monitors current ongoing processing and provides a signal to other brain regions when incompatible representations are active, including conflicts between automatic responses and current goals. This signal often indicates the need for additional processing from the regulatory control system. This later system implements more controlled processing to resolve conflict and direct processing in a goal-congruent fashion. The conflict-detection system is thought to be mediated by the anterior cingulate cortex (ACC) and the slower, regulatory-control system is thought to be mediated by regions of anterior and lateral prefrontal cortex (PFC). In the context of prejudice, automatic racial biases that contrast with egalitarian goals should trigger the conflict-detection system (Amodio et al., 2004; Amodio, Kubota, Harmon-Jones, & Devine, 2006), which would then recruit the regulatory-control system to modify biased processing.

This proposed relationship between automatic and controlled processing in racial bias was examined in a fMRI study (Cunningham et al., 2004). To isolate automatic and controlled processing of race, several egalitarian White participants were presented with Black and White faces for 30 ms or 525 ms based on the assumption that rapid subliminal presentation (i.e., 30 ms) would elicit automatic (and potentially unconscious) racial processing in regions like the amygdala, whereas the supraliminal presentation (i.e., 525 ms) would elicit relatively more controlled processing in the ACC and lateral PFC. As predicted, participants had greater amygdala activity following the subliminal Black than White faces. This differential amygdala activity to Black compared to White faces in the subliminal condition was highly correlated with individual differences in racial bias on the Implicit Association Test (Greenwald et al., 1998), replicating previous research (Phelps et al., 2000). In contrast, when the faces were presented supraliminally, this differential amygdala activity was significantly reduced, and brain regions involved in conflict-detection and regulatory-control (i.e., the ACC and lateral PFC) showed greater activity for Black compared to White faces. The authors then subtracted the Black-White difference in amygdala activity in the supraliminal condition from the corresponding difference in the subliminal condition to create an index of amygdala modulation. Participants with a high degree of amygdala modulation also showed the largest increases in ACC and lateral PFC activity in the supraliminal condition to Black compared to White faces.[5]

This pattern of results suggests that these egalitarian participants were controlling aspects of their automatic racial bias. Another study found a similar pattern, such that amygdala activity to Black faces was negatively correlated with lateral PFC (Lieberman et al., 2005). Further, the inverse

[5] The negative correlation between amygdala activity and ACC and lateral PFC activity to Black compared to White faces between the subliminal and supraliminal conditions occurred in a relatively simple perceptual task that was not explicitly focused on control. This raises a question about what exactly it is that White participants are to suppress or inhibit when they see a Black face, and why they would feel motivated to do so in a mere perceptual task (Amodio, 2008). This issue has actually been directly addressed by Richeson and colleagues (2003) in which differential engagement of the dlPFC during an almost identical task mediated the relationship between implicit measures of racial bias and impairment on a classic cognitive control task (the Stroop) following an interracial interaction. In addition, there is extensive evidence in the social psychological literature showing that people attempt to control their racial bias in a host of situations and tasks that do not explicitly require control, especially when people are motivated by personal beliefs or values to be egalitarian (see Crandall & Eshleman, 2003 for a review). Thus egalitarian participants may attempt to control emotional and cognitive responses to race, even when control is not explicitly required for the task.

relationship between the amygdala and lateral PFC to Black compared to White faces in the subliminal condition has recently been directly replicated using an index of functional connectivity, suggesting that this relationship is not limited to individual differences, but occurs within participants in an online fashion (Forbes, Cox, Schmader, & Ryan, 2010). Taken together, these studies have begun to identify the neural mechanisms involved in controlling automatic biases when participants have both the motivation and opportunity.

In the short term the controlled suppression of automatic biases can reduce the expression of these biases. However, controlled processing has a number of important limitations. There is extensive experimental evidence, for example, that controlled processes operate like a limited resource (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven & Baumeister, 2000) due to metabolic constraints (Gailliot & Baumeister, 2007). Specifically, attempts to suppress or override automatic or pre-potent responses reduce controlled resources. Thus participants with a large discrepancy between the strength of their automatic racial biases and their current goals may deplete their controlled resources faster than others, leading to the most behavioural discrimination during extended or sequential interracial interactions. Indeed, White participants with high levels of automatic racial bias on an IAT have the worst cognitive control on a Stroop task following an interracial interaction (Richeson & Shelton, 2003). Likewise, individual differences in implicit racial bias were positively correlated with PFC activity to Black (> White) faces during a simple dot detection task and cognitive impairment following an interracial interaction. Moreover, PFC activity to Black faces mediated the relationship between automatic racial bias and subsequent impairments in controlled processing (Richeson et al., 2003). Presumably participants with the most automatic racial bias had the most bias to control, and were therefore cognitively depleted.

The people who work the hardest to control their unwanted bias may ironically be the ones who suffer the costs, and may ultimately express these biases during sustained interactions. In a similar vein, efforts to suppress stereotypes and prejudice can rebound and actually increase the accessibility of these biases above baseline levels (Wegner, 1994). Moreover, the problems with top-down controlled processing are not limited to the mistreatment or perception of a target, but may lead to unhealthy physiological side effects (such as high blood pressure) (Gross, 1998; Gross & Thompson, 2007). However, controlling prejudice and stereotypes can be accomplished through other means. Instead of response-focused strategies aimed to suppress or inhibit an affective response, top-down antecedent-focused strategies can alter the initial activation of an affective response by construing the stimulus or context differently (Gross & Thompson, 2007).

Indeed, changing processing goals can alter the way low-level brain regions like the amygdala process positive and negative information about people (Cunningham, Van Bavel, & Johnsen, 2008). Antecedent-focused forms of control are often effective because they unleash a cascade of downstream effects on evaluation and behaviour.

Construing members of stigmatised social categories as individuals or as members of a different social category—an antecedent focused approach—may provide a powerful alternative to response-focused control. One processing goal that may be an especially powerful means of reducing bias is to individuate people and place less emphasis on their group membership (Brewer, 1988; Fiske & Neuberg, 1990). Individuating a member of a stigmatised group may reduce reliance on social category cues and the activation of stereotypes and prejudices. To test this hypothesis, a recent fMRI study had White participants process Black and White faces as individuals or as members of a social group (Wheeler & Fiske, 2005). Consistent with the research described above, when participants engaged in social categorisation (e.g., classifying the faces by age) they had greater amygdala activity to the Black than White faces. However, when participants were simply asked to consider the preferences of each individual (deciding whether each person preferred certain vegetables) they had greater amygdala activity to White than Black faces—reversing the standard amygdala response to race. These findings identify the neural substrates underlying the ability of participants to shape their own evaluative responses by attending to certain pieces of information and ignoring others.

As we noted above, people can have many dynamic and overlapping social identities, and their current self-categorisation with any of these identities can alter the way they learn and evaluate others (Tajfel, 1982; Turner et al., 1987; Turner et al., 1994). Any self-categorisation, whether deliberate or not, can presumably shift the construal of others in a manner congruent with the current salient identity. Accordingly, self-categorisation with any group—however minimal—should lead to the perception and evaluation of complex social stimuli in terms of their (minimal) group membership, ignoring other, orthogonal category dimensions. Building on this idea, we conducted a neuroimaging experiment to illustrate the role of self and social categorisation on amygdala function.

Similar to our other experiments, we randomly assigned White participants to a minimal mixed-race group and presented ingroup and outgroup faces to participants during neuroimaging (Van Bavel et al., 2008). Crossing race and group membership allowed us to examine the role of self-categorisation in neural processing; would membership in a novel group cause participants to process targets in terms of this novel group membership rather than race? Moreover, the mixed-race groups paradigm equated ingroup and outgroup members in familiarity and novelty.

Participants in the previous neuroimaging studies on race had different experiences and associations with the social categories, making it possible that differential novelty may have elicited differences in amygdala activity (Dubois et al., 1999). For example, White participants presented with White and Black faces would have likely been more familiar with White faces. We thought it was also possible that amygdala activity in the previous studies may have reflected a cognitive process other than simple valence (i.e., negativity or fear to Black faces). Research in our lab had previously shown that the amygdala may play a role in processing any motivationally relevant stimuli, regardless of valence (Cunningham et al., 2008). Thus we proposed that when race is the most salient social category, the amygdala may indeed be responsive to members of groups who are stereotypically associated with threat or novelty (Dubois et al., 1999). However, when race is not the most salient social category, groups that are currently relevant—minimal ingroup members in this case—would be associated with greater amygdala activity.

Participants were informed that they were in a study exploring learning about groups, that they had been assigned to the Leopards or Tigers, and that it was important for them to learn the members of their group and another group before moving to other phases of the study. Participants then learned the faces and responded to the faces during neuroimaging. On each trial, participants categorised one of the 24 faces in one of two ways: on explicit trials participants categorised each face according to team membership (Leopard or Tiger) and on implicit trials participants categorised each face according to skin colour (Black or White). These trials were referred to as explicit or implicit because participants' attention was explicitly focused on the current minimal group membership or not, respectively. As predicted, participants had greater amygdala activity to ingroup than outgroup faces (see Figure 7). Importantly, ingroup biases in neural processing occurred within minutes of team assignment, in the absence of explicit team-based rewards or punishments, and independent of pre-existing attitudes, stereotypes, or familiarity. Ingroup biases in neural activity were not moderated by target race or categorisation task, suggesting that they did not require explicit attention to team membership and may have occurred relatively automatically.

Whereas earlier studies often interpreted amygdala activity to outgroup faces as reflecting negativity or fear towards stigmatised group members, participants in our experiment (Van Bavel et al., 2008) had greater amygdala activity to ingroup members (see also Chiao et al., 2008). These results support the idea that the amygdala may be involved in segregating relevant from irrelevant stimuli in order to enhance perception of important stimuli (Anderson & Phelps, 2001; Vuilleumier, 2005; Whalen, 1998). The relevance of different social categories varies according to social context (Turner et al., 1987). In contexts where race provides the most salient group distinction, attitudes, cultural stereotypes, and personal values (egalitarianism) may provide the most relevant motivational guides. However, assigning people to mixed-race groups may change the way people construe race, and sensitise perceptual and evaluative processes to currently relevant group memberships. Indeed, people categorise others according to race when it is the salient social category, but categorise according to other group memberships (and ignore race) when they are part of a mixed-race group (Kurzban et al., 2001). Indeed, most previous neuroimaging studies make race the only salient difference between faces. The heightened amygdala activity to ingroup members in the current study may stem from their motivational relevance and salience in the current group context.

As we noted in the introduction, a fundamental assumption underlying the IR model is that brain systems are organised hierarchically, such that relatively automatic processes influence and are influenced by relatively more controlled processes in an iterative process. Whereas automatic processes provide relatively coarse perceptual and evaluative information, additional iterations allow for more controlled processes, which can interact with automatic processes and provide more nuanced or contextually appropriate evaluations. Building on these assumptions, we expected that information from the amygdala would be re-processed in higher-order regions to render a contextually appropriate evaluation (see Figure 2). For example, previous research has implicated the orbitofrontal cortex (OFC) in linking social and appetitive stimuli to hedonic experience (Kringelbach, 2005). This led us to predict that the OFC might not only be sensitive to positively valenced ingroup members (see also Volz, Kessler, & von Cramon, 2009), but also directly mediate the relationship between group membership and explicit measures of ingroup/outgroup preference.

To examine this relationship, we tested whether differential OFC activity to ingroup compared to outgroup members was associated with individual differences in self-reported ingroup bias (i.e., liking ingroup members more than outgroup members) (Van Bavel et al., 2008). As part of the experimental session described above, we had participants complete a computerised face-rating task *following* neuroimaging. Participants were told that ‘‘people can often quickly determine who they like or dislike based on subtle facial features and expressions’’ and asked to rate each of the 24 faces in random order on a 6-point liking scale ($1 =$ *dislike* to $6 =$ *like*). There were no group labels (i.e., Leopards and Tigers) on the screen during the rating task. Replicating previous research (see Brewer, 1979), participants reported a preference for ingroup compared to outgroup members (which was not moderated by race), such that participants reported liking ingroup members and were relatively neutral towards outgroup members. In other words, assigning participants to a group (the independent variable) caused them to report a relative preference for ingroup members compared to outgroup members (dependent variable): participants assigned to the Tigers reported liking members of the Tigers more, and the same ingroup preference was true of participants who were assigned to the Leopards. We created an individual difference index of self-reported ingroup bias (ingroup – outgroup) on the liking scale. Next we identified a region of the OFC (the mediator) that was more active to ingroup than outgroup faces. Again, we created an individual difference index of self-reported ingroup bias (ingroup – outgroup) on OFC activity. We examined the relationship between this evaluative preference and individual differences in ingroup bias in OFC activity (ingroup – outgroup).

As shown in Figure 8, participants who reported a stronger explicit evaluative preference for ingroup compared to outgroup members had relatively greater OFC activity to ingroup compared to outgroup members. Moreover, the effect of group assignment on self-reported liking was mediated by OFC activity, such that the effect of group membership on ingroup bias was significantly reduced when controlling for increases in OFC activity to ingroup compared to outgroup members. In other words, participants were assigned to one of two groups, which led to a relative increase in OFC activity for ingroup compared to outgroup members and led them to report a relative preference for ingroup compared to outgroup members. The effect of group membership on self-reported preferences for ingroup compared to outgroup members was mediated by OFC activity. Interestingly, this brain–behaviour relationship is similar to studies showing

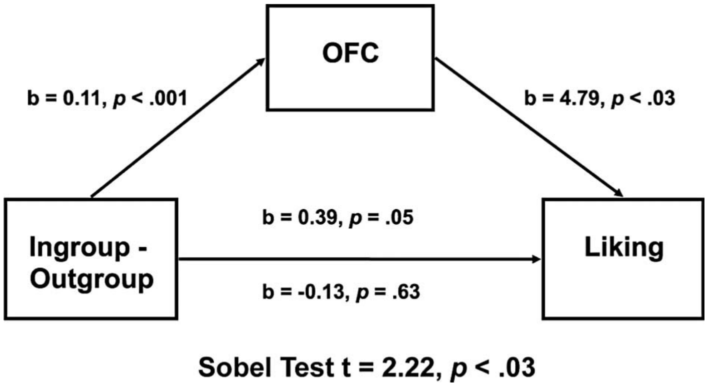

Figure 8. Participants were assigned to a group and responded to ingroup and outgroup faces during fMRI and then rated each face on a 6-point liking scale. This figure shows that the independent variable group membership (ingroup versus outgroup) has a statistically significant effect on the mediator orbitofrontal cortex (OFC) activity and the dependent variable self-reported liking, leading to ingroup bias (ingroup greater than outgroup) on both dependent measures. Specifically, participants had greater OFC activity to ingroup compared to outgroup members during fMRI and reported liking ingroup members more than outgroup members. Importantly, the effect of group membership (ingroup – outgroup) on self-reported liking was mediated by activity in the OFC.

a correlation between a similar region of the OFC and self-reported pleasantness of hedonic stimuli, such as food (e.g., Kringelbach, O'Doherty, Rolls, & Andrews, 2003). This region of the OFC appears to play a central role in representing and processing subjective value across stimulus domains (Kringelbach, 2005), including ingroup value.

SUMMARY AND FUTURE DIRECTIONS

The study of social perception has made substantial progress in the past half century, yet there remains considerable debate about the role of bottom-up and top-down influences on social categorisation. On the one hand, a host of social, cultural, and developmental factors continue to sow the seeds of prejudice. People inherit biases from family (Aboud, 1988), peers (Bagley & Verma, 1979), and the media (Jones, 1997) through an unrelenting stream of information about various social groups. Over a lifetime, constant exposure to stereotypes and prejudices generates deeply entrenched associations (Staats & Staats, 1958) that colour the way people see, feel, and act towards others. These automatically activated biases must be redressed through deliberate top-down control processes in which an initial reaction is consciously suppressed. On the other hand, there is evidence that emotionally or motivationally significant stimuli can alter perceptual processing (Balcetis & Dunning, 2006; Bruner, 1957; Vuilleumier, 2005) and the top-down influence of context and motivation has the potential to modulate a host of lower-order systems (Cunningham et al., 2008; Kim et al., 2004). We propose that automatic and ostensibly inevitable aspects of social categorisation are sensitive to this latter form of top-down influence by virtue of self-categorisation processes. In this section we revisit our proposed social neuroscience framework for self and social categorisation, discuss the implications of our research for reducing prejudice, link this work to multiple-categorisation and provide a brief summary.

Revisiting the social neuroscience approach to self and social categorisation

We proposed a social neuroscience approach to self and social categorisation that describes how representations associated with the contents of a currently salient self-categorisation can exert a top-down influence on automatic evaluation, social memory, and neural processing, overriding the effects of race—a visually and socially salient category. The current self-categorisation(s) is constructed from relatively stable identity representations stored in memory (such as the strength of one's social identity, one's role within the group, etc.) through iterative and interactive perceptual and evaluative processing. In this way the social context can make a particular social identity salient and elicit certain perceptual and evaluative processing consistent with the contents of that identity. Many of the experiments covered in this review employed variants of the minimal group paradigm to demonstrate the highly dynamic nature of self-categorisation and the implications of this process for social perception and evaluation. However, self-categorisation can reflect many different identities, and we propose that more important identities should generally exert an even greater influence on social perception and evaluation, but only when they are salient. The more important point is that our social identities can exert a profound influence on our representations of the social world, and override the influence of visually salient social categories.

We reviewed evidence from several behavioural and neuroimaging experiments showing the sensitivity of social perception and evaluation to the current self-categorisation, however minimal. Participants assigned to a minimal group evaluated and encoded targets on the basis of their minimal group membership (us vs them), rather than using another orthogonal category (i.e., race). The most arresting aspect of this research is that very brief exposure to these intergroup alliances was sufficient to elicit categorisation according to minimal group membership, making this a more potent social category than race, a category marked by years of exposure and associated with relatively stable stereotypes and attitudes. Apparently, racial categorisation may be malleable to a certain type of social context: one in which race is irrelevant to another psychologically salient social identity. Further, mere membership in an arbitrary group is sufficient to increase evaluative and behavioural preferences for ingroup members (Brewer, 1979); people who are actually *assigned* to one of the groups used group membership as a cue for categorisation rather than race, and revealed a preference for ingroup members, regardless of race, relative to those who were exposed to the groups but not a member in either one. Future research should examine whether this shift in perception and evaluation is moderated by other factors, including the extent to which people identify with their group and the nature of their role in the group.

We think it is important to clarify a number of issues related to the proposed social neuroscience framework of self and social categorisation. One might ask how this framework explains automatic racial bias, since White people do not necessarily walk around with chronically accessible racial identities. The proposed framework clearly predicts that contextually salient identities can override ostensible automatic racial biases. However, it does not claim that identity salience is *necessary* for the activation of social biases. Social categories may be salient due to visual properties, which may be one reason why biases based on race, gender and age are so pervasive, and this salience can be moderated by a variety of contextual factors other than identity. Another question is why the same social identity or social category will never be activated in an identical fashion twice. This is an untested assumption based on principles from computational neuroscience (O'Reilly & Munakata, 2000) and is presented in stark contrast to literature on the stability of stereotypes, evaluative biases and identities. However, our research strongly suggests that these biases are not always activated in the same way every time. Indeed, they may be overridden by very minimal intergroup distinctions (Van Bavel & Cunningham, 2009a; Van Bavel et al., 2008). We argue that exemplars and contexts shift in a dynamic fashion, and that the human mind is updated, however modestly, with new experience, such that the exact same exemplar, context, and state of the organism will almost never occur in the exact same combination on multiple occasions. Finally, people may argue that race can be positive or negative (depending on the context), but never irrelevant. Our framework and programme of research make the exact opposite claim: we argue that race can be made irrelevant in certain contexts. However, we are not suggesting that people are genuinely colourblind. It seems likely that race, like any physical or psychological property, may be represented in the brain, even when it is not exerting an influence on a specific mental process or task. Moreover, we believe that the context may not substantially change the underlying attitude representations stored in memory even when the current evaluations are radically altered. Thus, making race relevant—even in a mixed-race context—may quickly re-activate racial biases.

REDUCING PREJUDICE

He drew a circle that shut me out
Heretic, rebel, a thing to flout.
But Love and I had the wit to win:
We drew a circle that took him in!

Markham (1936, p.67)

The past half-century of research on prejudice has painted a troubling picture. Most North Americans appear to hold biases towards racial and other minority groups that are triggered by ordinary cognitive processes, like categorisation. Studies suggest that race affects perceptual processing within milliseconds (Ito & Urland, 2003) and appears to be highly salient and difficult to suppress (Park & Rothbart, 1982; Richeson & Shelton, 2003). Social psychologists have therefore suggested that encoding race may be inevitable, even when it is irrelevant to current perceptual goals (Hewstone, Hantzi, & Johnston, 1991; Stangor, Lynch, Duan, & Glass, 1992). It is therefore unsurprising that studies continue to reveal racial bias in real-world contexts, including discrimination in hiring (Bertrand &

Mullainathan, 2003) and juridical decision making (Rachlinski, Johnson, Wistrich, & Guthrie, 2009).

Social cognitive approaches to prejudice reduction have focused heavily on the response-focused suppression of race-based evaluations and stereotypes, inhibiting biases after they have been automatically activated (Monteith, Sherman, & Devine, 1998; Plant & Devine, 1998). However, evidence suggests that suppression is a narrow and inefficient form of regulation (Gross & Thompson, 2007) and can lead to increased stereotype accessibility (Macrae, Bodenhausen, Milne, & Jetten, 1994), cognitive depletion (Richeson & Shelton, 2003) or worse: unfriendly interracial interactions (Norton, Sommers, Apfelbaum, Pura, & Ariely, 2006). Our approach emphasises alternative, antecedent-focused routes to prejudice reduction: namely, changing the ways that others are construed (see also Amodio, 2010; Mendoza , Gollwitzer, & Amodio, 2010). This ‘‘new look’’ approach to changing automatic perceptual and/or evaluative processing may be especially important, because evaluation is dependent on information from early processes, and modest biases during the initial stages of perceptual and evaluation processing can have dramatic downstream effects (Cunningham et al., 2007). Although evaluation and behaviour may *feel* controlled and deliberate, initial automatic processes heavily inform both of them. We presented several experiments illustrating the power of self-categorisation to alter automatic components of social perception and evaluation and ultimately override ostensibly inevitable effects of race. Future research should examine whether this antecedent-focused approach to prejudice reduction bypasses the problems with suppression, including stereotype rebound and cognitive depletion.

By drawing a racially diverse circle—that is, re-categorising Black and White targets as members of an ingroup—participants generated positive evaluations of ingroup members, regardless of their race. Importantly, we found no evidence of outgroup derogation in any of our studies. Evidence that self-categorisation with a minimal mixed-race group elicits positive evaluations towards minimal ingroup members without eliciting negative evaluations towards outgroup members suggests that mixed-race groups may provide a socially constructive mechanism for attenuating racial bias. However, while automatic *evaluations* may be sensitive to the current salient self-categorisation, leading to a reduction in racial bias, the underlying prejudiced *attitudes* may remain relatively static (Cunningham et al., 2007). Thus, when participants are presented with Black and White faces who were unaffiliated with the ingroup and outgroup, they showed the standard pattern of racially biased evaluations (Van Bavel & Cunningham, 2009a). We also predict that these properties of the evaluative system will lead to the return of racial bias towards ingroup members when group membership is no longer psychologically salient. Thus, creating environments in which group membership is unrelated to existing social categories—keeping shared group memberships chronically salient—may be the best strategy for shifting evaluations away from more pervasive biases.

We have presented several studies showing that that minimal group categorisations can override race-based categorisations on a variety of indices, including implicit measures of racial bias. As we noted earlier, people who show evidence of racial bias on these measures tend to engage in discrimination in a variety of contexts (Dovidio et al., 2002; Greenwald et al., 2009; McConnell & Leibold, 2001), including hiring decisions towards racial and other minorities (Dovidio et al., 1997, 2002). An important question is whether the effects we have observed in our research extend to more real-world contexts. We generally assume that changes in evaluation will ultimately affect behaviour (Fishbein & Ajzen, 1975). We therefore expect that self-categorisation with a novel group should produce positive, pro-social interactions with ingroup members and reduce the discrimination that would otherwise be associated with race, at least among ingroup members. For example, simply assigning people to a game of pick-up basketball should swiftly change the perceptions, evaluations, and behaviour towards team-mates, leading to positive social interactions, regardless of race or creed. This may be especially true of the non-verbal forms of behaviour that characterise normal, spontaneous social interactions. According to the model presented here, we also believe categorisation is heavily context-dependent. As long as the context encourages the categorisation of others as ingroup or team members, people may have an easier time looking past race, gender or age. However, when the context makes these alternative, cross-cutting categories salient, patterns of perception, evaluation and behaviour should rapidly shift to reflect the current social reality. As a consequence, future prejudice-reduction interventions need to determine how to exploit and manipulate the effects of context in the real world.

It is also important to note that the social benefits of mixed-race group membership are offset by the caveat that self-categorisation leads to ingroup bias (Brewer, 1999). Although ingroup bias towards a minimal group may seem like a fair trade-off for racial biases laden with pernicious stereotypes, it is worth revisiting the effects of minimal group membership on more overt indices of intergroup discrimination (Tajfel et al., 1971). In many contexts, such as hiring or voting, any *differential* preference for one group over another may lead to the same pattern of behavioural discrimination, whether it is driven by ingroup bias or outgroup derogation. Therefore self-categorisation with a new group may offer a simple and promising approach to reduce racial bias but it must be carefully weighed against the possibility of spawning new forms of intergroup bias.

MULTIPLE CATEGORISATION

Crisp and Hewstone (2007) argue that there are two routes to reduced ingroup bias in crossed-category contexts: reducing intergroup differentiation and decategorisation. For example, the generation of a shared ingroup identity (e.g., University membership) that brings an outgroup on another social category dimension (e.g., Blacks) closer to the ingroup reduces racial differentiation and therefore reduces bias. Decategorisation occurs in more complex intergroup settings and involves a shift towards more individuated processing (Fiske & Neuberg, 1990). Although the current research was not designed to investigate these potential mechanisms, there appears to be evidence of both processes. Automatic evaluations towards Blacks were more positive when they were included in the ingroup, suggesting that the shared ingroup identity resulted in reduced differentiation and therefore reduced racial bias. However, these positive evaluations did not extend to unaffiliated Black faces (Van Bavel & Cunningham, 2009a). Thus it appears that the pro-social effects of the shared ingroup identity were constrained to the participants' mixed-race ingroup (i.e., their team). Participants assigned to a mixed-race group showed greater activity in the fusiform gyrus—a brain region involved in individuation—when they saw images of ingroup members (i.e., same team), regardless of race (Van Bavel et al., 2008). Taken together, these experiments raise the possibility that a shared ingroup identity may lead individuals to feel positive about *and* individuate ingroup members.[6]

Our research assumes that perception and evaluation are closely linked. However, several recent papers have found an effect of crossed categories on categorisation, but not evaluation (e.g., Vescio, Judd, & Kwan, 2004), leading researchers to question the link between social categorisation and intergroup bias (Park & Judd, 2005). We agree that social perception is not isomorphic with evaluative processing, and the dissociation between these processes may stem in part from additional component processes that alter evaluations to suit motivational concerns (Cunningham et al., 2007). However, the relationship between categorisation and evaluation will likely be closely linked when construals directly influence evaluation (or vice versa), but not when evaluations of social groups are prone to motivated suppression or inhibition (Crandall & Eshleman, 2003; Dunton & Fazio, 1997; Plant & Devine, 1998). Many of the experiments we reviewed address

[6] Note that categories, like race, may be difficult to ignore if they are central to an individual's self-definition. Thus, creating alternative, meaningful bases for categorisation in *conjunction* with existing categories like race may bypass the reactive effects of distinctiveness threat. Indeed, future research should examine whether individual differences in the centrality of race moderate the effects of our mixed-race manipulation on intergroup bias.

this issue by using measures of evaluation without explicit category labels or by presenting faces during perception and evaluation tasks without any visual cues to group membership. These experimental paradigms minimised the extraneous influence of specific category labels or visual cues on perception and evaluation and allowed multiple social categories (e.g., group and race) to spontaneously drive perception and evaluation. This approach revealed that participants spontaneously evaluated ingroup members according to their group membership.

A FINAL WORD

Our social and economic success hinges on our ability to cooperate with others from a variety of backgrounds. In a complex and dynamic social world, a central challenge for adaptive human behaviour is the flexible and appropriate categorisation and evaluation of others. The current paper takes a social neuroscience approach to self and social categorisation, linking the effects of self-categorisation and social identity on perception and evaluation to brain function. Our research illustrates that self-categorisation with a social group can dramatically shift social perception and evaluation, and override ostensibly pervasive racial biases. Although the effects of social categories such as race are relatively robust, our research shows that self-categorisation can alter the effects of race on variables ranging from perception to evaluation to brain function. Using a social neuroscience approach not only elucidates the neural substrates that implement self and social categorisation, it suggests that putatively hard-wired aspects of brain function are sensitive to the top-down influence of contextual and motivational factors.

REFERENCES

Aboud, F. E. (1988). *Children and prejudice*. New York: Blackwell.

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*, 469–479.

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

Amodio, D. M. (2008). The social neuroscience of intergroup relations. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 19, pp. 1–54). Hove, UK: Psychology Press.

Amodio, D. M. (2010). Coordinated roles of motivation and perception in the regulation of intergroup responses: Frontal cortical asymmetry effects on the P2 event-related potential and behaviour. *Journal of Cognitive Neuroscience*, *22*, 2609–2617.

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

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

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

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*, 305–309.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, *81*, 789–799.

Bagley, C., & Verma, G. (1979). *Racial prejudice: The individual and society*. Farnborough, UK: Saxon House.

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*, 612–625.

Bargh, J. A. (1984). Automatic and controlled processing of social information. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–41). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Bargh, J. A. (1994). *The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality & Social Psychology*, *62*, 893–912.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Transactions of the American Neurological Association*, *93*, 38–42.

Bernstein, M., Young, S., & Hugenberg, K. (2007). The cross-category effect: Mere social categorisation is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*, 709–712.

Bertrand, M., & Mullainathan, S. (2003). *Are Emily and Greg more employable than Lakisha and Jamal?: A field experiment on labor market discrimination. Working paper series WP 03-22*. Cambridge, MA: Massachusetts Institute of Technology Department of Economics.

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

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.

Brewer, M. B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*, 307–324.

Brewer, M. B. (1988). A dual process model of impression formation. In T. Srull & R. Wyer (Eds.), *Advances in social cognition* (Vol. 1). Hove, UK: Lawrence Erlbaum Associates Inc.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*, 429–444.

Brigham, J. C., & Ready, D. J. (2005). Own-race bias in lineup construction. *Law and Human Behaviour*, *9*, 415–424.

Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.

Bruce, V., & Humphreys, G. (1994). Recognising objects and faces. *Visual cognition*, *2/3*, 141–180.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*, 123–152.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organising factors in perception. *Journal of Abnormal and Social Psychology*, *42*, 33–44.

Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., & McClintock, M. K. (2000). Multilevel integrative analyses of human behaviour: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin*, *126*, 829–843.

Caruso, E., Mead, N., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National Academy of Sciences*, *106*, 20168–20173.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., et al. (2008). Cultural specificity in amygdala response to fear faces. *Journal of Cognitive Neuroscience*, *20*, 2167–2174.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

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

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 173–179.

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression of the expression and experience of prejudice. *Psychological Bulletin*, *129*, 414–446.

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorisation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 39, pp. 163–254). Orlando, FL: Academic Press.

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., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology*, *85*, 639–649.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*, 806–813.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170.

Cunningham, W. A., & Van Bavel, J. J. (2009). A neural analysis of intergroup perception and evaluation. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioural sciences* (pp. 975–984). Chichester, UK: Wiley.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science*, *19*, 152–160.

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*, 97–104.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multi-level framework for attitudes and evaluation. *Social Cognition*, *25*, 736–760.

Dasgupta, N., & Greenwald, A. G. (2001). Exposure to admired group members reduces automatic intergroup bias. *Journal of Personality and Social Psychology*, *81*, 800–814.

De Renzi, E., & Spinnler, H. (1966). Facial recognition in brain-damaged patients: An experimental approach. *Neurology 16*, 145–152.

Deschamps, J. C., & Doise, W. (Eds.). (1978). *Crossed category memberships in intergroup relations*. Cambridge, UK: Cambridge University Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

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

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

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*, 286–303.

Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., et al. (1999). Effect of familiarity on the processing of human faces. *NeuroImage*, *9*, 278–289.

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

Ellinwood, E. H. J. (1969). Perception of faces: disorders in organic and psychopathological states. *Psychiatric Quarterly*, *43*, 622–646.

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., Jackson, J. R., Dunton, B. C., & Williams, C. J. (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., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality & Social Psychology*, *50*, 229–238.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behaviour processes. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 97–116). New York: Guilford Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fiske, S. T., Lin, M. H., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. C. Y. Trope (Ed.), *Dual process theories in social psychology* (pp. 231–254). New York: Guilford.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York: Academic Press.

Forbes, C., Cox, C. L., Schmader, T., & Ryan, L. (2010). *Straight out of consciousness? Negative stereotype activation primes covariance between neural correlates of arousal, inhibition and cognitive control*. Unpublished manuscript, Tucson, Arizona.

Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of white racism, arousal and helping behaviour. *Journal of Personality and Social Psychology*, *35*, 691–707.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism*. Orlando, FL: Academic Press.

Gaertner, S. L., Rust, M. C., Dovidio, J. F., Bachman, B. A., & Anastasio, P. A. (1996). The contact hypothesis: The role of a common ingroup identity on reducing intergroup bias among majority and minority group members. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition?* (pp. 230–360). Newbury Park, CA: Sage.

Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review*, *11*, 303–327.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorisation in visual recognition studied with functional MRI. *Current Biology*, *7*, 645–651.

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

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise recognising novel objects. *Nature Neuroscience*, *2*, 568–573.

Gauthier, I., Tarr, M. J., Moylan, J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000). Does visual subordinate-level categorisation engage the functionally-defined face area? *Cognitive Neuropsychology*, *17*, 143–163.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.

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.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nature Neuroscience*, *2*, 574–580.

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

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

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

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*, 224–237.

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). New York: Guilford Press.

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: A comparison of faces and scenes. *Neuroimage*, *17*, 317–323.

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

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 versus ingroup face stimuli. *Neuroreport*, *11*, 2351–2355.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Hehman, E., Maniab, E. W., & Gaertner, S. L. (2010). Where the division lies: Common ingroup identity moderates the cross-race facial-recognition effect. *Journal of Experimental Social Psychology*, *46*, 445–448.

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorisation and person memory: The pervasiveness of race as an organising principle. *European Journal of Social Psychology*, *21*, 517–528.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in Social Categorisation: The role of prejudice and facial affect in race categorisation. *Psychological Science*, *15*, 342–345.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorisable individuals. *Journal of Personality and Social Psychology*, *85*, 616–626.

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, and Behavioural Neuroscience*, *5*, 21–36.

Ito, T. A., Willadsen-Jensen, E. C., & Correll, J. (2007). Social neuroscience and social perception: New perspectives on categorisation, prejudice, and stereotyping. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behaviour* (pp. 401–421). New York: Guilford Press.

James, W. (1896). *The will to believe and other essays in popular philosophy*. Cambridge, MA: University Press.

Johnson, M. K., & Hasher, L. (1987). Human learning and memory. *Annual Review of Psychology*, *38*, 631–668.

Jones, J. M. (1997). *Prejudice and racism*. (2nd ed.). New York, NY: McGraw Hill Companies, Inc.

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

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialised for the perception of faces. *Philosophical Transactions of the Royal Society of London B*, *361*, 2109–2128.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of fMRI responsivity to surprised faces. *Journal of Cognitive Neuroscience*, *16*, 1730–1745.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*, 623–634.

Krendl, A. C., Macrae, C. N., Kelley, W. M., & Heatherton, T. F. (2006). The good, the bad, and the ugly: An fMRI Investigation of the functional anatomic correlates of stigma. *Social Neuroscience*, *1*, 5–15.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*, 20600–20605.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience*, *6*, 691–702.

Kringelbach, M. L., O'Doherty, J., Rolls, E. T., & Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cerebral Cortex*, *13*, 1064–1071.

Kunda, Z., & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry*, *10*, 12–22.

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

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

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

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1364–1382.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*, 559–574.

Levine, R. A., & Campbell, D. T. (1972). *Ethnocentricism: Theories of conflict, ethnic attitudes and group behaviour*. New York: Wiley.

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.

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

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.

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

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*, 397–407.

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.

Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed.). New York: McGraw-Hill.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other "race". *Journal of Personality and Social Psychology*, *13*, 330–334.

Markham, E. (1936). Outwitted. In *The best loved poems of the American people* (p. 67). New York: Doubleday.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 734–760.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91–126). New York: Academic Press.

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.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3–35.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*, 512–523.

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

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers*, *36*, 630–633.

Mitchell, J. P., Nosek, B. N., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.

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

Mullen, B., Migdal, M. J., & Hewstone, M. (2001). Crossed categorisation versus simple categorization and intergroup evaluations: A meta-analysis. *European Journal of Social Psychology*, *31*, 721–736.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247–259.

Myrdal, G. (1944). *An American dilemma: The negro problem and modern democracy*. New York: Harper & Bros.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Ng, W-J., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, *25*, 217–232.

Norton, M. I., Sommers, S. R., Apfelbaum, E. P., Pura, N., & Ariely, D. (2006). Color blindness and interracial interaction: Playing the political correctness game. *Psychological Science*, *17*, 949–953.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dynamics*, *6*, 101–115.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, *56*, 717–734.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, *18*, 287–291.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*, 785–787.

Ostrom, T. M., & Sedikides, C. (1992). Outgroup homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*, 536–552.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the minimal group paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology*, *29*, 1049–1071.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*, 291–303.

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.

Park, B., & Rothbart, M. (1982). Perception of outgroup homogeneity and levels of social categorisation: Memory for the subordinate attributes of ingroup and outgroup members. *Journal of Personality and Social Psychology*, *42*, 1051–1068.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorisation and the process of intergroup bias. *Journal of Personality and Social Psychology*, *59*, 475–486.

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*, 57–75.

Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgement: Correcting for context-induced contrast. *Journal of Experimental Social Psychology*, *29*, 137–165.

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

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

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

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Proffitt, D. R. (2006). Distance perception. *Current Directions in Psychological Science*, *15*, 131–135.

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, *84*, 1195–1252.

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, *39*, 177–183.

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

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287–290.

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z-L., 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.

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.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*, 440–444.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, *115*, 15–36.

Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality and Social Psychology Bulletin*, *34*, 260–274.

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

Simmel, G., & Wolf, K. H. (1950). *The sociology of Georg Simmel*. Glencoe, IL: Free Press.

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

Sporer, S. L. (2001). Recognising faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*, 36–97.

Staats, A. M., & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, *57*, 37–40.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorisation of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*, 207–281.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behaviour. *Personality and Social Psychology Review*, *8*, 220–247.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*, 96–102.

Tajfel, H. (1972). La categorisation sociale (English trans.). In S. Moscovici (Ed.), *Introduction à la psychologie sociale*. Paris: Larouse.

Tajfel, H. (1982). *Social identity and intergroup behaviour*. Cambridge, UK: Cambridge University Press.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorisation and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level processing automized by expertise. *Nature Neuroscience*, *3*, 764–769.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorisation theory*. Oxford, UK: Basil Blackwell.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, *20*, 454–463.

Urban, L. M., & Miller, N. (1998). A theoretical analysis of crossed categorisation effects: A meta-analysis. *Journal of Personality and Social Psychology*, *74*, 894–908.

Van Bavel, J. J., & Cunningham, W. A. (2009a). Self-categorisation with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, *35*, 321–335.

Van Bavel, J. J., & Cunningham, W. A. (2009b). A social cognitive neuroscience approach to intergroup perception and evaluation. In W. P. Banks (Ed.), *Encyclopedia of consciousness* (pp. 379–388). New York: Academic Press.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of ingroup bias: A functional magnetic resonance imaging investigation. *Psychological Science*, *19*, 1131–1139.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2010). *Modulation of fusiform face area following minimal exposure to motivationally relevant faces*. Unpublished manuscript, Columbus, OH.

Vescio, T. K., Judd, C. M., & Kwan, V. S. Y. (2004). The crossed categorisation hypothesis: Evidence of reductions in the strength of categorisation but not intergroup categorisation but not intergroup bias. *Journal of Experimental Social Psychology*, *40*, 478–496.

Volz, K. G., Kessler, T., & von Cramon, D. Y. (2009). Ingroup as part of the self: Ingroup favouritism is mediated by medial prefrontal cortex activation. *Social Neuroscience*, *4*, 244–260.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594.

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.

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

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

Wilson, T. D., Samuel, L., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*, 1830–1839.