6

SOCIAL IDENTITY SHAPES SOCIAL PERCEPTION AND EVALUATION

Using Neuroimaging to Look Inside the Social Brain

Jay J. Van Bavel, Y. Jenny Xiao, and Leor M. Hackel

On February 26, 2012, 17-year-old Trayvon Martin was shot and killed by neighborhood watch captain George Zimmerman—who thought the teenager looked suspicious—sparking a national outrage and discussion about race in the United States. Why had an innocent, unarmed African American teenager died on his way home? In the aftermath, some sought answers not in explicit racism and prejudice, but rather in the domain of implicit racism—arguing, as two social psychologists wrote in an op-ed (newspaper opinion piece), that "our minds are colored by race" (Goff & Richardson, 2012, n.p.).

According to a standard dual process account developed over the last 25 years, the implicit effects of race that colored George Zimmerman's mind that night were sadly predictable. Seeing a Black face would likely trigger automatic negative evaluations and associations for Mr. Zimmerman, regardless of his explicit beliefs about race (see Devine, 1989). Moreover, due to his inexperience with Black faces, he may have failed to see Mr. Martin as an individual, but rather merely as an interchangeable member of the social category "African American." According to the standard dual process account, these events took place automatically and inevitably, triggered in a reflex-like manner by the salient visual cue of race. However, in this chapter, we provide evidence that the psychological and neural processes underlying person perception are context-dependent, dynamic, and shaped by motivational states, rather than inevitable, reflexive, and predictable. Specifically, we review research showing that *self-categorization* and *social identity* structure social perception and evaluation. Self-categorization involves the activation of psychological connections between the self and some class of stimuli at the personal (i.e., defining oneself as unique from others) or collective (i.e., defining oneself in terms of similar characteristics to one's social group) level (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987; Turner, Oakes, Haslam, & McGarty, 1994).

In the following pages, we will describe and critique the dual process account of racial bias, and present an alternative, dynamic systems approach. We will then review a series of experiments supporting this perspective by showing how self-categorization and social identity shape social perception and evaluation (Van Bavel & Cunningham, 2011). Specifically, we will present studies showing that social identities emerge very rapidly under minimal conditions, and can override biases in social perception and evaluation toward groups with much greater historical and cultural baggage, such as race. Importantly, however, we will present research showing that race is not merely "erased" from the visual system, suggesting that currently salient identities may guide low-level aspects of perception and evaluation even when aspects of race (e.g., physiognomic features) are still represented in the brain (see Kaul, Ratner, & Van Bavel, 2012; Ratner, Kaul, & Van Bavel, 2012). Finally, we will discuss the methodologies we used and comment on the value, costs, and benefits of these methods. Specifically, to examine the impact of self-categorization on ostensibly automatic components of the person perception network, we took a multi-level approach—termed social neuroscience (Cacioppo, Berntson, Sheridan, & McClintock, 2000). This approach is based on the assumption that complex phenomena, like prejudice, are best understood by combining social and biological theories and methods (see also Cunningham & Van Bavel, 2009; Van Bavel & Cunningham, 2009b). By understanding more about underlying neural processes, we hope to develop a functional understanding of the biological systems that underlie social perception and evaluation, and help to build the foundation for future research and intervention.

The Dual Process Approach: Race as a Bottom-up Cue in Social Perception

Our approach differs in a number of important ways from dual process and systems models, which have emerged and dominated various topic areas in psychology—including conceptualizations of stereotypes and prejudice (e.g., Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995)—since the 1980s. According to some dual process models, people have two systems for attitudinal representations: an *implicit* representational system, which gives rise to automatic attitudes; and an *explicit* system, which gives rise to controlled attitudes (e.g., Wilson, Samuel, & Schooler, 2000). This dual process approach was influential in early social neuroscience research on prejudice, as researchers sought to distinguish automatic (e.g., fast, unintentional, outside of awareness) and controlled (e.g., slow, deliberate) processes in the brain. For instance, research on this topic revealed that racial biases emerged within the first few hundred milliseconds of perceptual processing (Ito & Urland, 2003) and were linked to relatively early components in the face-processing network (Golby, Gabrieli, Chiao, & Eberhardt, 2001). Moreover, the neural substrates of the dual processes appeared to be separable, such that the amygdala, for example, was strongly correlated with *implicit*, but not *explicit*, measures of racial bias (Phelps et al., 2000). Meanwhile, people with sufficient motivation and opportunity to control biased responses appeared to employ the anterior cingulate cortex (Amodio et al., 2004) and lateral prefrontal cortex (Cunningham et al., 2004), regions that have both been implicated in cognitive control and self-regulation, to help control these biases (see Figure 6.1).

A Dynamic Systems Approach: Identity Shapes Social Perception and Evaluation

Although dual process models serve as a useful heuristic for the human mind and have sparked numerous studies, advances in social and cognitive neuroscience suggest that the human evaluative system is more widely distributed and dynamic than that proposed by traditional dual process models (see Van Bavel, Xiao, & Cunningham, 2012). As such, we argue that there is unlikely to be a clear dissociation between explicit and implicit representations in the brain (see Cunningham, Zelazo, Packer, & Van Bavel, 2007; Van Bavel, Xiao et al., 2012). Rather, we argue that representations are constructed from the dynamic interaction of multiple brain systems, and feature the recursive interaction between bottom-up cues (e.g., skin color or hair length) and top-down cues (e.g., attention or motivations) that interact in cycles until the evaluative system settles on a

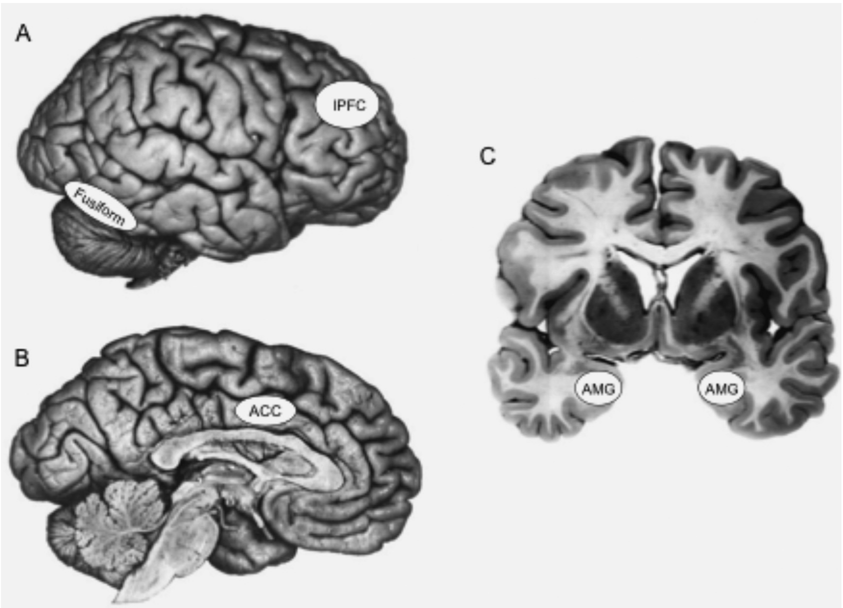

FIGURE 6.1 Anatomical Images of the Human Brain Identifying the Approximate Spatial Locations of the (A) Fusiform Gyri and Lateral Prefrontal Cortex (lPFC), (B) Anterior Cingulate Cortex (ACC), and (C) Left and Right amygdala (AMG)

representation of a target (Cunningham et al., 2007; Freeman & Ambady, 2011). Further, the dynamic approach suggests that virtually all aspects of evaluation and perception states are influenced by the context and motivation of the perceiver. As such, we assume that early aspects of social perception depend on social identity more than on the ostensibly automatic processes assumed by most dual process models of personal perception and evaluation (see Van Bavel & Cunningham, 2011 for an extended discussion).

Therefore, we argue that salient social identities—rather than race, per se—can shape person perception processes in this dynamic fashion. It is true that biases and stereotypes about certain social groups, such as racial and gender groups, are often built upon years of exposure to cultural stereotypes and personal experience. However, the more general process of categorizing one's self as a group member (i.e., self-categorization) can have an early influence on intergroup perception, evaluation, and behaviors (Turner et al., 1994; Van Bavel & Cunningham, 2011). As social beings, humans are remarkably adept at dividing up the world into *us* and *them*, even in the absence of any factors typically posited to account for intergroup bias, such as intergroup competition or conflict (Tajfel, Billig, Bundy, & Flament, 1971). In a classic minimal groups paradigm used by social identity researchers, people form groups quickly and favor members of their own group, even when groups are formed under rather arbitrary premises (e.g., the flip of a coin). Such evidence points to the value humans place on social identities and the context-dependent nature of identification. In fact, self-categorization with a group may occur in a reflexive and automatic manner, and guide automatic evaluations and behaviors (e.g., Otten & Wentura, 1999). Building on self-categorization theory, we assume that self-categorization is "inherently variable, fluid, and context dependent" (Turner et al., 1994, p. 454). In other words, the social context can heighten the accessibility of a particular social identity (e.g., a racial identity, a national identity, etc.), which in turn elicits perceptions and evaluations consistent with the activated aspects of this identity.

Furthermore, in terms of the time course of these processes, we argue that prior states of the evaluative system set the stage for automatic construals of subsequent stimuli. Higher-order processes—mediated by top-down control signals from the frontal and parietal networks—can incorporate expectations, goals, bodily states, and context into representations that are deemed most relevant in a given context (see Miller & Cohen, 2001), which can then lead to different patterns of self- and social categorization. When a given self-categorization becomes salient, one may in turn be more likely to see others in light of their membership vis-à-vis this salient category. As a consequence, this may pre-empt ostensibly automatic racial bias (or other visually salient social categories) that some researchers have characterized as inevitable. Thus, the preceding context and motivational state of an organism informs subsequent evaluative processes (and vice versa) in a continually dynamic manner. Based on this model, we reasoned that self-categorization with a group should shape people's perceptions and evaluations of others in terms of their current and most salient self-categorization, not necessarily race. We used a social neuroscience approach to test whether the neural underpinnings of social categorization more closely match the proposed dynamic model of the human mind than dual process models (see Van Bavel & Cunningham, 2011, for a more detailed discussion).

How We Tested Our Perspective

With only behavioral measures, it may be difficult to illustrate *when* in the processing stream higher-order constructs such as self-categorization and social identity exert influence on perception and evaluation, and it would be impossible to demonstrate *where* in the brain these influences occur. By tapping into neural activity with functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), we could more precisely compare different models of social processing, and delineate differences between alternative models of perception and evaluation.

Functional magnetic resonance imaging provides an indirect index of neural activity by measuring changes in cerebral blood flow, which correlates with neural activity (see Huettel, Song, & McCarthy, 2004, for a review). Compared to other non-invasive measures of neural activity, fMRI offers superior *spatial* resolution (on the order of cubic millimeters), although it has inferior *temporal* resolution to methods like EEG (on the order of 4–6 seconds for fMRI, as opposed to milliseconds [ms] for EEG; see Cunningham, Packer, Kesek, & Van Bavel, 2009, for a discussion). As such, fMRI is better suited to answer questions about the localization of cognitive processes, whereas EEG is better suited to delineate their time course. Examining the neural substrates of person perception allowed us to test the impact of social categorization along the neuroaxis. Does self-categorization merely alter activity in brain regions involved in reflective, controlled processing (e.g., lateral prefrontal cortex)? Or does it also affect activity in brain regions implicated in ostensibly bottom-up, automatic processing (e.g., the amygdala)?

Empirical Evidence: Dissociating the Effects of Race and Self-categorization

Behavioral Investigations

To examine the relationship between self-categorization and intergroup perception and evaluation, we conducted several experiments in which we assigned participants to one of two mixed-race groups (e.g., the Lions or Tigers) with an equal number of Black and White males in each group. Participants then had a few minutes to memorize the group membership of these faces before we assessed dependent measures of their attitudes (Van Bavel & Cunningham, 2009a), memory

(Van Bavel & Cunningham, 2012b; Van Bavel, Swencionis, O'Connor, & Cunningham, 2012), and/or brain activity (Van Bavel, Packer, & Cunningham, 2008, 2011). Importantly, the faces on each team were fully counterbalanced across participants to ensure that any effects of group membership were due to self-categorization and identification, and *not* the visual properties of different faces (e.g., attractiveness, luminance, symmetry, etc). Moreover, race was orthogonal to team membership and there were no visual cues to group membership during the administration of the dependent measures. This mixed-race paradigm allowed us to examine whether individuals automatically and inevitably categorize others according to a visually salient category like race (e.g., Devine, 1989; Ito & Urland, 2005). We predicted that a salient self-categorization—however minimal—would override or even pre-empt racial bias.

In a pair of initial studies, we examined the influence of a shared social identity on ostensibly automatic evaluations (Van Bavel & Cunningham, 2009a). We measured automatic evaluations of the faces described above using a computerized response-window priming task (Draine & Greenwald, 1998). During this task, participants were instructed to categorize a word rapidly on each trial as "good/liked" or "bad/disliked." On each trial, a face appeared at the center of the computer monitor for 150 ms (followed by a blank screen for 50 ms) before an unambiguously positive (e.g., love) or negative (e.g., hatred) target word appeared. We assumed that faces with positive associations would increase accuracy to positive words and decrease accuracy to negative words. By the same logic, faces with negative associations would decrease accuracy to positive words and increase accuracy to negative words.

Replicating previous research, 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 (Fazio et al., 1995). In other words, mere exposure to a racially diverse environment was not sufficient to override racial bias. In contrast, participants who were actually part of a mixed-race group had positive automatic evaluations of White *and* Black in-group members, and these evaluative preferences were driven by in-group favoritism and not out-group derogation. That is, group membership increased relative positivity toward Black in-group members relative to Black out-group members, eliminating the standard pattern of automatic racial bias among in-group members. Thus, participants' evaluations reflected their current salient self-categorization even when the groups had no history of contact or conflict, and when there was an orthogonal, visually salient social category cue (i.e., race) with strong existing evaluative connotations. In short, "automatic" racial bias was *not* automatic.

Neuroimaging Investigations

Although the neural networks involved in evaluation are widely distributed (Cunningham et al., 2007), initial research focused on a small structure in the temporal lobe called the amygdala (see Figure 6.1 above), a region in the extended face network that plays an important role in social evaluation (see Macrae & Quadflieg, 2010). The amygdala has been implicated in a host of social and affective processes, including fear conditioning and processing of negative stimuli (for a review, see Phelps, 2006). More strikingly, the amygdala is activated even during rapid subliminal presentations of affectively significant faces (Whalen et al., 1998). Several studies have found that individual differences in amygdala activity for Black compared to White faces correlate with implicit measures of racial bias including the Implicit Association Test (IAT) (Cunningham et al., 2004; Phelps et al., 2000). These correlations with racial bias led some researchers to interpret differences in amygdala activation in intergroup contexts as evidence of negativity toward stigmatized groups. However, differences in amygdala activity to faces of different racial groups are generally uncorrelated with explicit measures of prejudice (Phelps et al., 2000).

The dissociation between implicit and explicit measures of racial bias, along with the dissociation between explicit measures of racial bias and amygdala activity, are consistent with numerous dual process models of prejudice. Indeed, several studies have examined the control of automatic racial biases. For example, one fMRI study examined both automatic and controlled responses to Black and White faces (Cunningham et al., 2004). Several White participants were presented with Black and White faces for 30 ms or 525 ms. Consistent with the assumption that rapid subliminal presentation (i.e., 30 ms) would elicit automatic racial biases, participants had greater amygdala activity following exposure to subliminal Black faces than following subliminal White faces. Moreover, differential amygdala activity in response to Black versus White faces was correlated with individual differences in racial bias on the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998). In contrast, when the faces were presented supraliminally (i.e., 525 ms), this differential amygdala activity was significantly reduced, and brain regions involved in conflict detection and regulatory control (i.e., the anterior cingulate cortex [ACC] and lateral prefrontal cortex [PFC]) showed greater activity for Black compared to White faces. Such findings suggest that participants were controlling their automatic racial bias.

Based on our research on the malleability of automatic evaluations (Van Bavel & Cunningham, 2009a), however, we reasoned that flexibly construing people as in-group members might provide a powerful alternative to the traditional dual process models of control evident in previous research (see also Wheeler & Fiske, 2005). We had previously shown that the amygdala may play a role in processing any motivationally relevant stimuli, regardless of valence (Cunningham, Van Bavel, & Johnsen, 2008). Thus, we reasoned 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—explaining some previous findings from neuroimaging studies of prejudice. In contrast, when race is not the most salient social category, groups that are currently motivationally relevant would be associated with greater amygdala activity. In minimal group contexts, in-group members tend to be motivationally relevant and afford group members the opportunity to meet belonging goals (Van Bavel & Cunningham, 2012b; Van Bavel, Swencionis et al., 2012).

Similar to our other experiments, we randomly assigned White participants to a minimal mixed-race group, asked them to learn the faces of each group, and then presented them with in-group and out-group faces during neuroimaging (Van Bavel et al., 2008). Crossing race and group membership allowed us to examine the role of self-categorization in neural processing: Would membership in a new group lead participants to process targets in terms of this salient group membership rather than race? Importantly, assigning people to mixed-race groups equated in-group and out-group members in familiarity and novelty. Participants in previous neuroimaging studies on race had different experiences and associations with Black versus White social categories, making it possible that novelty with Black faces may have elicited differences in amygdala activity (Dubois et al., 1999).

In our experiment, during neuroimaging, participants categorized each face according to either team membership (Leopards or Tigers) or skin color (Black or White). As predicted, participants had greater amygdala activity to in-group (i.e., same-team) than out-group (i.e., other-team) faces, regardless of task. Importantly, this in-group bias in neural processing occurred within minutes of group assignment, in the absence of explicit team-based rewards or punishments, *and* independent of pre-existing attitudes, stereotypes, or familiarity. In-group bias in neural activity was not moderated by target race or categorization task (i.e., categorizing by team or by skin color), suggesting that it did not require explicit attention to group membership and may have occurred relatively automatically. Again, this suggests that self-categorization can shape relatively automatic aspects of social perception and evaluation.

Whereas earlier studies often interpreted amygdala activity to out-group faces as reflecting negativity or fear toward stigmatized group members, participants in our experiment (Van Bavel et al., 2008) had greater amygdala activity to in-group members. These results support the idea that the amygdala may be involved in segregating relevant from irrelevant stimuli to enhance perception of *important* stimuli (Anderson & Phelps, 2001; Vuilleumier, 2005; Whalen, 1998). Importantly, 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, racial attitudes, cultural stereotypes, and personal values (e.g., egalitarianism) may provide the most relevant motivational guides. Most neuroimaging studies in this literature make race the only salient difference between faces by showing 50% White faces and 50% Black faces. This departure from the population base-rate where most participants live may artificially make race and racial stereotypes more salient than usual and therefore heighten racial biases reflected in amygdala activity in those studies. However, assigning people to mixed-race groups may change the way people construe race and other social categories, and sensitize perceptual and evaluative processes to other contextually relevant group memberships. Indeed, people categorize others according to race when it is the most salient social category, but categorize according to other group memberships (and ignore race) when they are part of a mixed-race group (Kurzban, Tooby, & Cosmides, 2001). The heightened amygdala activity to in-group members in the current study may stem from their motivational relevance and salience in the current group context.

Empirical Evidence: Own Race Bias or Own "Group" Bias?

Next, to obtain more evidence that self-categorization, rather than more bottom-up aspects of race, drives social perception, we extended our research to one of the most robust and widely replicated phenomena in social perception—the own-race bias. Extensive research has shown 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., http://www.alllooksame.com). Although the ORB may appear to be relatively innocuous, it can lead an eyewitness in a criminal case to misidentify a suspect from another race, leading to the conviction of an innocent person. Indeed, approximately 36% of wrongful convictions are due to erroneous cross-race eyewitness identification in which Caucasian witnesses misidentify minority defendants (Scheck, Neufeld, & Dwyer, 2000).

For the past several decades, perceptual expertise—a bottom-up mechanism—has been widely accepted as the primary psychological explanation for ORB. According to this account, people become expert at identifying individuals within their own race by virtue of greater exposure to own-race individuals, including family, friends, and acquaintances, relative to members of another race. This increased exposure produces a specific expertise for encoding and/or recalling own-race faces. Over the course of a lifetime of interactions with people from the same race, experience in making both within- and between-race distinctions tunes the perceptual system to make finer distinctions among exemplars within own-race faces than within other-race faces (Malpass & Kravitz, 1969).

At the neural level, one fMRI study (Golby et al., 2001) examined the relationship between the ORB and activation in the fusiform face area (FFA), a sub-region of the fusiform gyrus (see Figure 6.1, above) located on the ventral surfaces of the temporal lobe (Kanwisher, McDermott, & Chun, 1997). Building on research showing that FFA activity increases with visual expertise (see Palmeri & Gauthier, 2004, for a review), Golby and colleagues (Golby et al., 2001) presented Black and White participants with pictures of Black and White faces during fMRI. Activity in FFA was greater to own-race than other-race faces for both Black and White participants (see also Lieberman, Hariri, Jarcho, Eisenberger, & 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 in-group members. Consistent with the "perceptual expertise" hypothesis, 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).

More recently, social categorization approaches have challenged the perceptual expertise model of ORB (Hugenberg, Young, Bernstein, & Sacco, 2010; Sporer, 2001). According to these approaches, categorizing others as in-group or out-group members may alter the depth or type of processing they receive, such that own-race faces are processed as individuals by default and other-race faces as interchangeable representatives of a social category, leading to superior recognition memory for own-race faces (Bernstein, Young, & Hugenberg, 2007; Sporer, 2001). Moreover, activity in the fusiform may not be fully contingent on expertise with specific categories, but rather may be sensitive to top-down factors, such as the level of categorization that perceivers use to recognize stimuli (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997).

In line with this top-down approach and our previous work, we predicted that people assigned to a minimal group would encode in-group members at a subordinate level and out-group members at a superordinate level, and that this differential encoding would be reflected in differences in fusiform activity (in-group > out-group), despite participants' limited exposure to members of both categories. We reasoned that if the fusiform is merely processing expert stimuli, then White participants would show greater fusiform activity to White faces relative to Black faces, regardless of contextual information. In contrast, if the fusiform is flexibly involved in individuating stimuli—which is subject to the influence of categorizing another person as an in-group member (Bernstein et al., 2007; Brewer, 1988)—participants would show greater activity to in-group relative to out-group faces, regardless of race. Indeed, consistent with the latter hypothesis, we found greater activation within the bilateral fusiform gyri for in-group than out-group faces (Van Bavel et al., 2008). These results provide convergent evidence that the fusiform is sensitive to shifts in social contexts, responding selectively to face stimuli imbued with psychological significance by virtue of their group membership and encoding the more motivationally relevant in-group 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; Van Bavel & Cunningham, 2012b).

We have recently replicated this pattern of in-group bias in the face-sensitive sub-region of the fusiform (i.e., the FFA) and shown that relatively greater activity in this region mediates the effects of group membership on recognition memory—

a behavioral index of individuation (Van Bavel et al., 2011). Specifically, we found a positive correlation between the FFA differences for in-group versus out-group faces and recognition memory differences for in-group versus out-group faces. These findings imply that in-group members are more likely to be processed as individuals or exemplars in a non-categorical fashion than out-group members, consistent with social cognitive models of person perception (Brewer, 1988; Fiske & Neuberg, 1990; Sporer, 2001). Thus, our research provides evidence that the motivational relevance of categories, like group membership, can affect fusiform activity in a flexible and dynamic fashion even in the absence of explicit task instructions or long-term experience with the category.

We believe our study suggests that the fusiform may play a key role in processing in-group members in greater depth than out-group members—placing in-group biases in perception firmly within the realm of motivated social perception (Balcetis & Dunning, 2006). As such, these studies on the fusiform, along with a series of recent behavioral studies (see Hugenberg et al., 2010, for a review), led us to predict that social memory would also be sensitive to the motivational aspects of social identity. In a series of studies, we examined the influence of social identification, social roles, and belonging needs on recognition memory (Van Bavel & Cunningham, 2012b; Van Bavel, Swencionis et al., 2012). As predicted, we found that the motivational aspects of the perceiver's social identity shape social attention and memory over and above mere categorization into groups. For instance, participants who reported a strong need to belong or were highly identified with their minimal in-group had a memory advantage for in-group versus out-group faces. However, social affordances attenuated in-group bias: Memory for out-group faces was heightened among participants who were assigned to a role (i.e., spy) that required attention toward out-group members. This research suggests that many aspects of social identity play a role in shaping social perception.

Although we have now accumulated extensive evidence that social identity can override or pre-empt racial bias in a number of domains, we are not suggesting that people are becoming colorblind to race. It seems likely that race may be represented in the brain, even when it is silent on a specific mental process or task, simply because it generally co-varies with visually salient features (e.g., color). Indeed, we have recent evidence that race may be encoded in the visual system, even when it does not produce racial biases (Ratner et al., 2012). As we noted above, our previous work suggests that overall activation levels in the fusiform vary as a function of salient group membership, not necessarily race. However, we re-analyzed the data using multivariate pattern analysis—a technique that can identify *patterns* of neural activity representing a type of stimulus even in the absence of greater *mean activation* in response to that stimulus category. We found that patterns of neural activity within the early visual cortex and a face-sensitive sub-region of the fusiform gyri (FG) represented the race of faces, even though the FG showed similar *overall* levels of activation to White and Black faces (Kaul et al., 2012;

Ratner et al., 2012). Moreover, race was represented in the fusiform to a greater extent than in the early visual cortex, suggesting that fusiform activity did not merely reflect low-level perceptual information (e.g., color) from the early visual cortex. The results indicate that *patterns* of activation within the FG encode race even when the *mean level* of fusiform activation is driven by other group memberships. Therefore, the human visual system may still encode color and physiognomic features that allow them to distinguish between Black and White faces, even when other more motivationally salient social categorizations pre-empt or override their influence on specific processes or tasks like face individuation. This also suggests that racial bias may (re-)emerge relatively quickly if race is made psychologically salient (see Van Bavel & Cunningham, 2011, for a discussion).

Implications

Over the past few decades, dual process and systems models have emerged as the dominant perspective in psychology (Chaiken & Trope, 1999). In particular, the development and widespread utilization of implicit measures, buttressed by research on automaticity, have suggested that stereotypes and prejudice can operate automatically and efficiently below conscious control and despite intentions to the contrary (Greenwald & Banaji, 1995). Take racial attitudes, for example: Many dual process models would predict that exposure to a Black target *automatically* gives rise to negative associations in White participants, and non-prejudiced perceivers are motivated to then exert *control* over these attitudes (Devine, 1989).

In contrast to dual process models, our research takes a dynamic systems approach to social perception and evaluation (Van Bavel, Xiao et al., 2012). Using a combination of social cognitive and neuroscience techniques, we show the flexible nature of ostensibly automatic social perception, as well as the mechanism through which these influences occur. We demonstrate that even rapidly and arbitrarily created social categories, which people have no prior knowledge about or contact with, can pre-empt or override the influence of existing social categories that often carry a great deal of societal and historical baggage, such as racial groups. Specifically, we show that automatic evaluations of and neural responses to other people are shaped by salient social and self-categorization.

Perhaps the most arresting aspect of this research is that very brief exposure to arbitrary intergroup alliances was sufficient to elicit categorization and identification 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. Therefore, racial categorization may be malleable in certain contexts, including ones in which race is irrelevant to another psychologically salient social identity. Further, mere membership in an arbitrary group is sufficient to increase evaluative and behavioral preferences for in-group members; people who are actually *assigned* to one of the groups used group membership as a cue for categorization rather than race, and revealed a preference for in-group members, regardless of race, relative to those who were exposed to the groups but not made a member of either one.

Thus, while visually salient categories like race may trigger bottom-up, perceptual processes due to low-level visual features (e.g., physiognomic features), a psychologically salient social identity can trigger top-down perceptual and evaluative processing, which can attenuate the ostensibly automatic effects of race. Moreover, 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-categorization can pre-empt or override the effects of visually salient and socially important categories—including categories with which people have extensive experience—perhaps before these social categories even begin to influence the perceptual and evaluative system.

This chapter is a sample of a broader research program examining the influence of social identity on perceptions of the *social* world, including rapid attention orientation (Brosch & Van Bavel, 2012), person memory (Van Bavel & Cunningham, 2012b; Van Bavel, Swencionis et al., 2012), automatic evaluation (Van Bavel & Cunningham, 2009a), and empathy (Cikara, Bruneau, Van Bavel, & Saxe, 2013). More strikingly, we have also recently found that social identity can shape group members' perception of the *physical* world, including their perception and representation of physical distance (Xiao & Van Bavel, 2012) and olfaction (Coppin et al., 2013).

Costs and Benefits

In a similar way to other chapters in this volume, we use a combination of techniques and methodologies from social psychology and cognitive neuroscience. Were it not for the development and utilization of techniques like fMRI, investigation of the flexibility of perception and evaluation would have remained on the level of behavior. For instance, in a behavioral study, we could manipulate presentation durations of our stimuli and use reaction-time measures to look at the time course of processing social stimuli. However, with neuroscience techniques, we can examine the sensitivity of different brain regions to top-down influences, challenging hard-wired views of brain function. Taking social perception, for example, we now know that the evaluative function of the amygdala is not driven purely in a bottom-up fashion by low-level features of stimuli, but also by higher-order motives (Cunningham et al., 2008; Van Bavel et al., 2008). This empirical evidence would have been impossible to establish without utilization of neuroimaging techniques to supplement our existing body of literature, which has a long history of behavioral investigations.

Moreover, our approach offers new perspectives to social psychology and cognitive neuroscience. By bringing a neuroscience approach to social psychology, we can explore concrete mechanisms underlying the abstract process of self-categorization and the construct of social identity. Additionally, considering a neuroscience perspective may help us to re-evaluate classic psychological theories, challenging the distinction between automatic and controlled processing, and suggesting that a strict dissociation is unlikely to be a natural kind grounded in brain structure or function. Meanwhile, by bringing social psychological theories to cognitive neuroscience, this research demonstrates the flexibility of the person perception network and the power of "top-down" systems to alter supposedly "bottom-up" processing. The social cognitive neuroscience perspective highlights the power of the situation to influence cognition, suggesting that even basic cognitive processes must be considered within the social identity contexts in which they are situated (Turner et al., 1994). For example, processing in the fusiform gyri—part of the ventral visual stream—appears to depend on the current motivational relevance of the stimulus being processed.

On a broader level, this work implies a model of human cognition in which a currently active mindset determines supposedly "automatic" reactions. Top-down influences on cognition and emotion need not emerge only after unavoidable automatic responses, with these responses being controlled through strategies like suppression or reappraisal. Rather, top-down influences before the fact—what we term "pre-appraisal" (Van Bavel, Xiao et al., 2012)—can determine which responses become active in the first place, thus influencing cognitive, evaluative, and affective outcomes.

Although we focus on fMRI research in this chapter, it is not well suited for addressing all types of research questions (see Cunningham et al., 2009 for a discussion). For instance, fMRI is usually more suited to delineating activities in distinct brain regions and different sub-areas, while EEG may directly examine the time course of psychological processes such as automaticity (Cunningham et al., 2009). Therefore, we have conducted several studies using EEG to examine the time course—a central feature of automaticity—of these processes. These studies suggest that social identity and other motivational factors can shape responses to faces within 100 ms of face presentation (Cunningham, Van Bavel, Arbuckle, Packer, & Waggoner, 2012; Van Bavel & Cunningham, 2012a). In other words, social identity concerns may *pre-empt* the activation of ostensibly automatic racial bias (see also Van Bavel & Cunningham, 2009a).

On a theoretical level, it is important to recognize that as useful as these methodologies are, they are by no means suited to all types of psychological research questions. In our lab, we spend a considerable amount of time trying to identify the optimal methodology and level of analysis for each research question. Ultimately, understanding social perception and evaluation with multiple methods and across multiple levels of analysis offers the greatest promise of generating integrative, process-oriented theories of self- and social categorization, and subsequently developing interventions for social issues like prejudice and discrimination.

Despite these numerous benefits, it is crucial to recognize the potential pitfalls of applying neuroscience techniques to investigations of social psychological questions, as well as some potential disadvantages in comparison with behavioral assessments. On the theoretical level, the use of reverse inference, in which one infers the presence of a cognitive process based solely on brain activity, is potentially problematic—as one brain region may be involved in a number of psychological processes—and researchers should exercise caution in this respect (see Poldrack, 2006). Researchers may want to include behavioral manipulations and measures that link cognitive processes to brain activity, and consider whether prior research has established a clear and sufficiently exclusive link between a particular psychological process and a particular brain region to support a reverse inference.

Furthermore, in terms of practicality, there are several potential disadvantages of using neuroscience techniques in social psychology research. First, research involving neuroscience techniques is usually much more costly than behavioral investigations. Second, facilities suited to conducting neuroscience research may be less readily accessible. Moreover, it usually requires prolonged training to acquire sufficient expertise to utilize neuroscience techniques and conduct appropriate data analysis. Although these practical concerns may be more mundane than potential theoretical pitfalls, it is nevertheless important to keep these in mind when conducting social neuroscience research.

Conclusion

In a complex and dynamic social world, a central challenge for adaptive human behavior is the flexible and appropriate categorization and evaluation of others. In this chapter, we discuss a social neuroscience approach to self- and social categorization, linking the effects of self-categorization and social identity on perception and evaluation to brain function. Our research illustrates that self-categorization with a social group can dramatically shape social perception and evaluation, and can pre-empt or override ostensibly pervasive racial biases. Although the effects of social categories such as race are relatively robust, our research shows that self-categorization can alter the effects of race on variables ranging from perception to evaluation, including underlying brain function. Using a social neuroscience approach not only helps to elucidate the neural substrates that underlie self- and social categorization, but also suggests that even putatively hard-wired aspects of brain function are sensitive to the top-down influence of contextual and motivational factors. As such, our responses to someone like Trayvon Martin may be influenced by seemingly trivial features of our salient social identities—such as whether we define ourselves as members of a neighborhood watch or a community association—-rather than being hard-wired and predetermined.

Suggestions for Further Reading

Theoretical

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.

This article provides a theoretical framework for understanding how attitudes and evaluations can emerge dynamically over time, with contextual and motivational influences, as an alternative to strict dual process frameworks.

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

This article outlines self-categorization theory, according to which representations of self- and social categorization are flexibly constructed and determine social perception.

Van Bavel, J. J., & Cunningham, W. A. (2011). A social neuroscience approach to self and social categorisation: A new look at an old issue. *European Review of Social Psychology, 21*, 237–284.

This article integrates research on self-categorization and social neuroscience.

Methodological

Amodio, D. M. (2010). Can neuroscience advance social psychological theory? Social neuroscience for the behavioral social psychologist. *Social Cognition, 28*, 695–716.

This article discusses the use of neuroscience methods in social psychological research, including when it can be useful and, importantly, how to avoid mistakes and pitfalls in social neuroscience.

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

A guidebook to theory and practice in the use of fMRI.

Ochsner, K. N. (2007). Social cognitive neuroscience: Historical development, core principles, and future promise. In A. Kruglanski & E. T. Higgins (Eds.), *Social psychology: A handbook of basic principles* (Vol. 2, pp. 39–66). New York: Guilford Press.

This chapter presents a history, rationale, and statement of aims of social cognitive neuroscience.

References

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.

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

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.

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

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brosch, T., & Van Bavel, J. J. (2012). The flexibility of emotional attention: Accessible social identities guide rapid attentional orienting. *Cognition, 125,* 309–316.

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

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

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2013). The roots of intergroup empathy bias: Intergroup competition and differentiation shape empathy toward in-group and out-group members. Unpublished manuscript, MIT.

Coppin, G., Delplanque, S., Oud, B., Margot, C., Sander, D., & Van Bavel, J. J. (2013). Swiss identity smells like chocolate: Social identity shapes olfactory experience. Unpublished manuscript, University of Geneva.

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., Packer, D. J., Kesek, A., & Van Bavel, J. J. (2009). Implicit measures of attitudes: A physiological approach. In R. E. Petty, R. H. Fazio & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 485–512). New York: Psychology Press.

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 behavioral sciences* (pp. 975–984). Hoboken, NJ: Wiley.

Cunningham, W. A., Van Bavel, J. J., Arbuckle, N. L., Packer, D. J., & Waggoner, A. S. (2012). Rapid social perception is flexible: Approach and avoidance motivational states shape P100 responses to other-race faces. *Frontiers in Human Neuroscience, 6*, 140.

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., 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.

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

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.

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.

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.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*, 247–279.

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

Goff, P. A., & Richardson, L. S. (2012). Running from race in our minds. *Huffington Post*. Retrieved from http://www.huffingtonpost.com/phillip-atiba-goff/trayvon-martin-race_b_1376621.html

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.

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.

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

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The Categorization-Individuation Model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*, 1168–1187.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable 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 Behavioral Neuroscience, 5*, 21–36.

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

Kaul, C., Ratner, K. G., & Van Bavel, J. J. (2012). Dynamic representation of race: Processing goals shape race encoding in the fusiform gyri. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nss138

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

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.

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., pp. 428–463). 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.

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

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.

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., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729–738.

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

Ratner, K. G., Kaul, C., & Van Bavel, J. J. (2012). Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nss063

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.

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

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization 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.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization 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.

Van Bavel, J. J., & Cunningham, W. A. (2009a). Self-categorization 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., & Cunningham, W. A. (2011). A social neuroscience approach to self and social categorisation: A new look at an old issue. *European Review of Social Psychology, 21*, 237–284.

Van Bavel, J. J., & Cunningham, W. A. (2012a). The dynamic nature of social perception in the human brain. Unpublished manuscript, New York University.

Van Bavel, J. J., & Cunningham, W. A. (2012b). A social identity approach to person memory: Group membership, collective identification, and social role shape attention and memory. *Personality and Social Psychology Bulletin, 38,* 1566–1578.

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

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: Evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience, 23*, 3343–3354.

Van Bavel, J. J., Swencionis, J. K., O'Connor, R. C., & Cunningham, W. A. (2012). Motivated social memory: Belonging needs moderate the own-group bias in face recognition. *Journal of Experimental Social Psychology, 48*, 707–713.

Van Bavel, J. J., Xiao, Y. J., & Cunningham, W. A. (2012). Evaluation as a dynamic process: Moving beyond dual system models. *Social and Personality Psychology Compass, 6*, 438–454.

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

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.

Xiao, Y. J., & Van Bavel, J. J. (2012). See your friends close, and your enemies closer: Social identity and identity threat shape the representation of physical distance. *Personality and Social Psychology Bulletin, 38*, 959–972.