

CHAPTER 50: A NEURAL ANALYSIS OF INTERGROUP PERCEPTION AND EVALUATION

William A. Cunningham and Jay J. Van Bavel

In an era of increasing globalization, social and economic harmony depends on the ability of people to cooperate with others from a variety of ethnic, geographic, and religious backgrounds. A trend toward explicitly egalitarian attitudes among North Americans has been accompanied (and motivated) by legislation that makes discrimination a crime and public scrutiny that makes a single racist statement a major political liability. Yet, although a majority of Americans now report nonprejudiced attitudes and strong motivations to respond without prejudice, investigations over the past few decades have shown that the majority still have lingering automatic and perhaps unconsciously activated negative responses toward many minorities and socially disadvantaged groups (Nosek, Banaji, & Greenwald, 2002). These subtle prejudices have been shown to activate even among individuals with egalitarian motivations, and appear to take considerable cognitive effort to control once released. These prejudices have also been shown to directly predict discrimination, including negative nonverbal behavior and biased hiring decisions toward racial and other social groups (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). In a recent meta-analysis, these automatic forms of prejudice were stronger predictors of discrimination than self-report when there was a strong desire to hide or conceal one's attitude (Greenwald, Poehlman, Uhlmann, & Banaji, in press). To deal with these challenges, basic research is needed to understand the structures and mechanisms that promote social prejudice: from large societal structures and historical events, to the genetic and neural mechanisms that provide the basic machinery humans use to understand and navigate their complex social worlds.

Although research examining the neural bases of prejudice using neuroscience methods has a relatively short history, this research has already made dramatic progress. Looking across several key studies, one conclusion can be easily drawn—the processing of social group membership (typically investigated in a racial context) appears to influence nearly all aspects of brain function from early visual processing to higher order aspects of executive function and deliberate thought. The widely

distributed patterns of brain activity found to covary with the processing of social groups suggest that the simple categorization of people into groups influences not a single group perception system, but rather a constellation of processes that collectively give rise to a multitude of social biases. Although the processes underlying social perception require further specification, initial neuroscience research on social prejudice provides important hints for our understanding of the mechanisms of prejudice and intergroup discrimination.

Social Cognitive Perspective

The study of prejudice has been at the forefront of social psychology for over a half century. Ever since Allport (1954) wrote his classic book, *The Nature of Prejudice*, and placed research on prejudice firmly within mainstream social psychology, psychologists have sought to understand prejudice and find effective means for its elimination. For obvious reasons, much of this work studied overt acts of discrimination and verbally reported statements of prejudiced attitudes. However, evidence shows that people may also spontaneously evaluate social objects along a good-bad dimension, without necessarily being aware that they are doing so (Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Given such findings, models of social attitudes suggest at least two modes of evaluation: one that involves conscious and controlled modes of thinking and another that involves relatively automatic processes that operate without deliberate thought and sometimes without conscious awareness (Greenwald & Banaji, 1995; Nisbett & Wilson, 1977). Importantly, an evaluation following more controlled processing may differ from an evaluation based only on more automatic processing. On indirect or implicit measures that tap automatic associations, many White participants show negativity toward Blacks, the elderly, or foreigners compared with Whites, the young, or Americans, respectively; yet they report unbiased attitudes on questionnaires that allow more controlled or conscious evaluations of the same groups (Cunningham, Nezlek, & Banaji, 2004; Devine, 1989; Nosek et al., 2002). When the social context discourages expressions of prejudice, automatic biases can be a stronger predictor of discrimination than self-report measures, particularly of subtle, nonverbal acts (Greenwald et al., in press).

Evidence that prejudice can operate automatically has led researchers to the troubling conclusion that verbal reports are not always bona fide indicators of prejudice. If people cannot fully report on the ways

that prejudice influences their thoughts, feelings, and behaviors, then the extent to which group biases permeate social cognition may be underestimated. With this in mind, it is necessary to examine the ways that prejudice, both in its conscious and unconscious forms, influences each step of social perception, all the way from early visual processing where social categories are initially encoded and applied, to the reflective processing used to generate more or less biased perceptions as a function of high-order goals and values. Using the classic computer metaphor in cognitive science in which the mind processes information in a serial sequence of processing stages, this chapter examines how social categories shape each stage of social perception. Information processing at each stage is dependant on information outputted from preceding stages, which implies that small biases occurring during initial stages may have dramatic downstream effects. Thus, although later consciously accessible evaluative processes feel as if they are under our deliberate control, they can be heavily biased by automatic forms of prejudice that influence processing and behavior prior to conscious reflection.

Figure 50.1

Note. (Yellow) Fusiform gyrus: Involved in visual perception and recognition. A subregion of the fusiform, known as the fusiform face area, has been shown to be particularly active to presentations of faces, and face-like stimuli. Beyond just faces, this region appears to play a role in making distinctions within categories of stimuli (e.g., cars), especially among experts. (Orange) Amygdala: A small structure in the medial temporal lobe that plays a role in the encoding and processing of affective representations. Activation in the amygdala is commonly found following the presentation of affectively intense stimuli (e.g., fear), although a more general role for the processing of any motivationally significant stimulus has been proposed. (Purple) Insula: A region located within the somatosensory cortex. The anterior insula, in particular, receives direct input about homeostatic and visceral information from the body and sends output to other limbic (including amygdala) and cortical regions. The insula has been linked to the experience of disgust and other emotional states. (Blue) Medial prefrontal cortex: The medial region of the anterior frontal lobes. The medial PFC has been implicated in social and affective processes, including self-referential processing and simulating the mental states of other (termed mentalizing). (Red) Anterior cingulate: A functionally heterogeneous region of the cingulate cortex. The anterior ACC—especially the dorsal region—appears to play a key role in monitoring for cognitive conflict. (Green) Dorsolateral prefrontal cortex: The lateral regions of the anterior frontal lobes. The lateral PFC appears to play an important role in cognitive control and executive function, including the processes involved in working memory.

“They All Look Alike to Me”: Perception and Categorization

The unfortunate, yet too often overheard, phrase in the title of this section underscores that prejudices can influence the very way that individuals see the world; and specifically, that people appear to be better at processing and remembering people from their own race than from other races—an effect that has been termed the *same-race bias* (Malpass & Kravitz, 1969). Although the same-race bias may seem relatively harmless at first glance, it can have serious implications for crucial decisions in the real world. With eyewitness testimony being among the most compelling pieces of evidence in criminal trials, the misidentification of a suspect from another race can literally lead to a death sentence for an innocent person (Brigham & Ready, 2005), especially when paired with certain societal stereotypes and prejudices (e.g., Seeleman, 1940).

If people are less likely to identify outgroup relative to ingroup members, it is possible that this bias occurs in the early stages of visual processing—outgroup faces may not be processed at the same level of detail as ingroup faces. One brain region that has been proposed as particularly important for perceptual biases in social processing is an area of visual cortex known as the fusiform gyrus (see Figure 50.1). In particular, the fusiform face area (FFA), has been shown to respond preferentially to faces (as contrasted with almost any other type of visual stimulus) (Kanwisher, McDermott, & Chun, 1997) and involves in the extraction of low-level perceptual features that can allow for individuation. To examine the role of the FFA in the same-race effect Golby, Gabrieli, Chiao, and Eberhardt (2001) used fMRI to examine the brain regions associated with facial processing while participants viewed same-race and other-race faces, as well as objects (radios). Black and White participants viewed pictures of Black and White faces to compare the processing of ingroup and outgroup members. As expected, the FFA was more sensitive to ingroup than outgroup 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 (superior memory for same-race over other-race faces) was predicted by fusiform gyrus activation to racial ingroup members at encoding.

Although this research provides an important link between early perceptual processing and racial biases in memory, these data are silent with regard to the psychological mechanism(s) that give rise to the difference (Levin, 2000; Sporer, 2001). On the one hand, it is possible that

the same-race bias is the result of familiarity with same-race faces. According to this view, people have a lifetime of experience interacting with family, friends, and acquaintances of the same race, and consequently become experts at automatically processing and distinguishing members of their race. As such, the bias is not necessarily motivational, but rather the consequence of long accrued perceptual experience. On the other hand, it is also possible that this bias is the result of motivated social perception (Balcetis & Dunning, 2006); categorizing others as ingroup or outgroup members may alter the depth of processing that they receive. People might view ingroup members as more important and be more likely to process them as individuals, in contrast to less relevant outgroup members who are lumped together (even perceptually) simply as “them.” Thus, while a lifetime of greater experience with one’s own groups may help give rise to or enhance the same-race bias, the simple categorization of others into an ingroup or outgroup may be *sufficient* to generate biases in intergroup perception and memory (Bernstein, Young, & Hugenberg, 2007).

A recent study tested these competing hypotheses by randomly assigning participants to one of two novel groups: the Leopards or Tigers (Van Bavel, Packer, & Cunningham, 2008). After participants learned the members of each group, they were presented with the same faces during fMRI. Importantly, participants had equal prescanning exposure to the ingroup and outgroup faces. To the extent that differences in group processing in the fusiform gyrus occur simply because of perceptual familiarity, no differences to ingroup and outgroup faces should be found. In contrast, if fusiform activity is sensitive to motivated aspects of social perception, including current self-categorization, then we should expect greater fusiform activity to ingroup as opposed to outgroup faces. Results supported the second hypothesis: the fusiform gyrus was more sensitive to novel ingroup than outgroup faces. This study suggests that the individuation of ingroup members as opposed to outgroup members may begin at the earliest stages of information processing and that this differentiation can be driven by the simple classification of others into groups.

Together, these data are consistent with the idea that ingroup members may be processed at a more individuated level than outgroup members even at the earliest stages of information processing (Rhodes, Byatt, Michie, & Puce, 2004). Whereas ingroup members are processed as individuals, extracting information about what makes each person unique, outgroup members are processed as interchangeable members of a general social category (see also Outgroup Homogeneity Effect; Ostrom & Sedikides, 1992). As such, outgroup members are more likely to be stereotyped, and these stereotypes are less likely to be disconfirmed by

individuating information. Consistent with this idea and complementing the individuation of ingroup members, people are faster to categorize other-race faces according to their race than own-race faces—an effect that has been labeled the *other-race* categorization advantage (Levin, 1996; Valentine & Endo, 1992). Using event-related potentials (ERPs), Caldara, Rossion, Bovet, and Hauert (2004) showed that the brain response to categorizing other-race faces (Asian faces) was about 20 ms faster than for own-race faces. Remarkably, these effects were seen a mere 240 ms after stimulus presentation, providing strong evidence for rapid and automatic differences in the processing and categorization of social groups.

These studies indicate the extent to which race and group membership influence early aspects of social perception. From an information processing perspective, initial perceptual processes can influence subsequent processes and ultimately lead to discrimination and injustice. If group membership influences the way that we see the social world and how we unconsciously divide others into meaningful categories, it should not be surprising that these early processes are going to affect downstream evaluations and behavior.

Affective Evaluations and Prejudice

A central focus of the neuroscience research on intergroup relations has been prejudice—the (typically negative) affective response people have to social groups as a whole, as well as individual members of those groups. Although the neural networks involved in an affective evaluative response are likely diffuse (Cunningham & Zelazo, 2007; Cunningham, Zelazo, Packer, & Van Bavel, 2007), initial research began with a focus on the amygdala. The amygdala is a small structure in the temporal lobe linked to an array of social and affective processes, including learning emotional information (Phelps, 2006), perceiving emotional faces (Whalen et al., 1998), and directing attention to important stimuli (Vuilleumier, 2005). More directly, with its tight connection to fear conditioning, threat processing, and negative affect more generally (LeDoux, 1986; Phelps, 2006), the amygdala was a logical starting place to investigate social prejudice. Importantly for the study of automatic affective biases, amygdala activation to negative emotional expressions (e.g., greater to fearful than to neutral facial expressions) has been found to be similar whether stimuli are presented at durations long enough for the stimuli to be consciously seen

(Morris et al., 1996) or more briefly (33 ms and masked) (Whalen et al., 1998). This suggests that the amygdala may play a critical role in rapid and unconscious evaluation of the environment.

In the first fMRI study of prejudice, Hart and colleagues (2000) showed Black and White participants blocks of Black and White faces. While this initial study revealed greater amygdala activation to outgroup than to ingroup faces (White participants viewing Black faces and Black participants viewing White faces), these results were qualified by a small sample size ($N = 8$) and relatively weak effects (the reported pattern was only observed in the second half of the study). Armed with a larger sample size, Lieberman, Hariri, Jarcho, Eisenberger, and Bookheimer (2005) replicated these results for White participants, but found the opposite pattern for Black participants, who also showed greater amygdala activation to Black than White faces. Despite the discrepancies between these initial studies, they provided evidence of a link between the processing of social group membership and a subcortical (perhaps automatic and unconscious) affective response.

To directly investigate the relationship between prejudiced attitudes and amygdala activity, subsequent research examined the relationship between amygdala activation and behavioral measures of prejudice. Phelps and colleagues (Phelps et al., 2000) presented White participants with Black and White faces while their amygdala was scanned during fMRI. Following the scanning procedure, participants completed both indirect (reaction time and physiological) and direct (self-report) measures of prejudice. If the amygdala was involved in prejudice, it was hypothesized that more prejudiced participants would show greater amygdala response to Black than White faces. Further, because the amygdala has been shown to be involved in unconscious evaluation and there tends to be greater variability on automatic/indirect than self-report measures of prejudice, it was hypothesized that a stronger relationship would be observed for the indirect than the direct measures. Although the study did not find overall greater amygdala activation to Black than White faces, both the reaction time measure (the Implicit Association Test; Greenwald, McGhee, & Schwartz, 1998) and the physiological measure (startle eyeblink) were significantly correlated with more amygdala activity to Black than White faces. Further, the explicit measure of prejudice was uncorrelated with amygdala activity.

Although Phelps et al. provided evidence for an important link between individual differences in automatic prejudice and amygdala activation, several important questions remained. Most importantly, if automatic evaluative biases are so pervasive, why were there no main effects of amygdala activation? In a follow-up study, Cunningham and

colleagues (2004) reasoned that paradigms used in previous fMRI studies of race bias may confound multiple processes and may obscure the full power of unconscious bias. Specifically, because long blocks of Black and White faces were presented supraliminally, participants may automatically evaluate the Black faces more negatively than White faces, but may also then try to control or suppress their responses (discussed later in this chapter). To more closely examine unconscious race bias, the participants were presented with Black and White faces subliminally (for 30ms), so that only automatic, unconscious processes could be used to evaluate the stimuli. Further confirming the role of the amygdala in the automatic evaluative processing of social groups, significantly greater amygdala activity was found for subliminal Black than White faces in all participants but one. In addition, this differential amygdala activation was correlated ($r = .79$) with the Implicit Association Test, the same indirect measure of prejudice used by Phelps and colleagues (2000).

Although these studies link the amygdala to automatic racial bias, the exact role the amygdala plays in prejudice remains unclear. For example, a patient with bilateral amygdala damage has shown racial bias on the IAT, demonstrating that the amygdala is not necessary for the expression of automatic prejudice (Phelps, Cannistraci, & Cunningham, 2003). Moreover, a valence-specific conceptualization of amygdala activation has been called into question by studies showing that positive as well as negative stimuli both evoke amygdala activity (Hamann, Ely, Hoffman, & Kilts, 2002). Further research will be critical in determining the role of amygdala when processing social group membership. These studies not only will lead to a better understanding of prejudice, but will also aid in our understanding of amygdala function per se.

Power of the Ingroup

Perhaps because negative visceral aspects of prejudice are the most frightening and salient to observers, dislike of outgroups has received much more attention than the reciprocal form of prejudice—positive associations toward ingroup members. However, the history of intergroup conflict provides strong evidence that ingroup love is a more common root of discrimination than “outgroup hate” (Brewer, 1999). Moreover, in contexts where discrimination arises as a result of differential evaluations of two groups, ingroup bias can lead to the same patterns of discrimination as outgroup derogation (e.g., in the context of a hiring decision, ingroup bias and outgroup derogation would both lead a White candidate to be hired over a Black candidate). Although these decisions are the result of quite

different affective processes, the result is identical—a Black candidate is treated unfairly and the cycle of discrimination continues.

Recent research has begun to dissociate the neural processes involved in ingroup and outgroup biases. In one study, participants were asked to think about the opinions and preferences of a person who had a similar or dissimilar political affiliation (Mitchell, Macrae, & Banaji, 2006). In the current partisan political landscape in the United States, it was assumed that more politically identified participants would process the similar person as an ingroup member, and therefore activate brain areas that have been linked to self-referential processing. Liberals were expected to be more motivated to understand the mental states of another liberal than a conservative, and the converse was expected for conservatives. This is exactly what was found. Considering the mental state of a similar other lead to activity in ventral areas of medial prefrontal cortex (PFC), whereas considering the mental state of a dissimilar other lead to activity in more dorsal areas of medial PFC. Interestingly, individuals who strongly self-categorized with a political group, as measured by an implicit measure, had greater *ventral* medial activity to politically similar others and less *dorsal* medial PFC activity to dissimilar others. Because regions of medial PFC have previously been implicated in building mental models of other minds and simulating the thoughts and feelings of other people (called mentalizing, Mitchell, 2006), with more ventral areas being more involved in the processing of self-relevant information (Kelley et al., 2002), the authors concluded that although similar and dissimilar others both recruit regions involved in understanding others, similar others were more likely to be processed like the self.

If people are more willing or able to mentalize about people with whom they share a group membership, especially a group identity they highly value, certain outgroup members may not receive this processing. Consistent with this hypothesis, Harris and Fiske (2006) found that when participants viewed members of social outgroups that typically arouse feelings of contempt, such as drug users, less ventral medial prefrontal cortex activation was observed. Instead, for these stigmatized group members, the insula—a brain area associated with the emotion of disgust (Phillips et al., 1997)—showed more activation. This pattern of data is consistent with the idea that not only can negative emotions be activated in response to outgroup members, but there may be certain aspects of prejudice marked by less processing for outgroups compared with ingroups. People may use less mentalization for certain groups of people over others (Cortes, Demoulin, Rodriguez, Rodriguez, & Leyens, 2005; Vaes, Paladino, Castelli, Leyens, & Giovanazzi, 2003).

Undoing the Automatic: The Deliberate Control of Prejudice

With a large body of research demonstrating that people are evaluated as members of social groups automatically, unconsciously, and sometimes unfairly, one might be inclined to take a pessimistic view of human nature. More optimistically, however, research has also documented that, at least under some circumstances, people can control automatic responses, and sometime even replace evaluations of one valence (a negative affective response) with another (a positive affective response). Among the more cherished aspects of human cognition is its ability to use controlled processing and abstraction to escape immediate stimulus-response contingencies and generate more nuanced evaluations and judgments in the service of long-term goals and values (Cunningham & Zelazo, 2007; Greenwald & Banaji, 1995). Behavioral research provides evidence for this suggestion, showing that when people have the motivation and opportunity to use more deliberate forms of cognitive processing, the influence of automatically activated stereotypes and prejudice is dramatically reduced (Devine, 1989; Dovidio et al., 1997; Fazio, 1990; Fazio, Jackson, Dunton, & Williams, 1995). Thus, although initial intergroup categorization and evaluation have important implications for intergroup relations, human behavior is often driven by values, goals, and motivations.¹

Social cognitive studies of prejudice regulation has tended to focus on the inhibition or suppression of initial evaluations deemed inappropriate or suboptimal (Devine, 1989; Petty & Wegener, 1993). In this view, the automatic activation of prejudiced representations and biased processing leads to discriminatory behavior unless controlled intervention eliminates these biases. For this to be successful, two sets of processes are thought to be necessary—a *conflict-detection system* and a *regulatory control system*—each with different temporal dynamics and neural generators (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Cohen, Botvinick, & Carter, 2000; MacDonald, Cohen, Stenger, & Carter, 2000). The conflict-detection system automatically monitors current representations and provides a signal that additional processing resources are required when incompatible representations are active. In the case of an egalitarian person, prejudiced representations that contrast with egalitarian goals would trigger this conflict detection system, which may then recruit the regulatory control system to update and modify prejudiced representations. The *conflict-detection* system is thought to be mediated by the anterior cingulate gyrus,

and the slower, more reflective *regulation* system is thought to be mediated by regions of anterior and lateral PFC.

A study by Cunningham and colleagues (2004) provided evidence that these regulatory systems play an important role in modulating automatic affective responses to race. As noted in the section on evaluation, Cunningham and colleagues presented White participants with Black and White faces for 30 or 525 ms. Although these participants reported having strong egalitarian values, they also showed more automatically activated negative responses to the social category Black than White on an Implicit Association Test (IAT) (Greenwald et al., 1998). Showing that automatic prejudice operates unconsciously as well as that people can control this response White participants had greater amygdala activation to Black than to White faces (which were randomly intermixed), but only when the faces were presented subliminally (30 ms), such that participants did not report seeing the faces. In contrast, Black faces in the supraliminal (525 ms) condition were associated with activity in brain regions involved in controlled processing and executive function, such as the anterior cingulate cortex and lateral PFC. Moreover, the reduction in amygdala response during the supraliminal condition was inversely correlated with activity in these areas of anterior cingulate cortex and dorsolateral PFC. This pattern was consistent with the idea that people with the motivation and opportunity can control their automatic responses to social groups.

Providing further evidence that the PFC is involved in the regulation of prejudice and building on the idea that explicit linguistic processing generally inhibits affective processing (Lieberman, 2003), Lieberman and colleagues (2005) asked both Black and White participants to categorize Black and White faces according to linguistic labels ("African American" versus "Caucasian") or perceptual information (e.g., matching one Black face to another Black face). When participants categorized faces according to perceptual information, they had greater amygdala activity to Black than White faces. In contrast, there were no differences in amygdala activity when individuals categorized according to linguistic labels. Similar to the study by Cunningham and colleagues (2004), this lack of amygdala difference was accompanied by heightened lateral PFC to the Black than White faces, and this increase in lateral PFC activity was associated with the decrease in amygdala activation to Black faces. Again, these data show that the PFC is involved in modulating presumably more automatic responses under certain conditions.

Although these fMRI studies implicate the PFC in the regulation of prejudiced responses and show that these areas can decrease differential responses to Black and White faces in the amygdala, they are silent to how quickly these processes unfold. To study the temporal aspects of the

conflict-detection system in prejudice control, Amodio and colleagues (2004) measured ERPs while participants categorized rapidly presented images as tools or guns. Immediately preceding each object, a Black or White face appeared (see Payne, 2001 for more details). The task was designed so that if a Black face automatically activated concepts of negativity (or the specific stereotype of violence), then participants would be more likely to make errors misidentifying a tool as a gun following the presentation of a Black face. For egalitarian participants, these errors should activate the conflict-detection system because their behavior (prejudiced response) would be incongruent with their values and ideals. As predicted, prejudicial errors among egalitarian participants were followed by an ERP signal that has been previously associated with the anterior cingulate in general, and the conflict detection system specifically—the ERN (error-related negativity). Importantly, the ERN in this study occurred within 200 ms of response errors, providing evidence that people do automatically monitor for unconscious prejudice and may be able to trigger corrective processes relatively automatically.

Although deliberate control is typically associated with inhibition, reflective aspects of emotion regulation can also be involved in the maintenance or enhancement of affective states to construct a more intense explicit evaluation. Many social groups evoke negative affective responses without an increase in compunction or guilt (Crandall, Eshleman, & O'Brien, 2002). Few people attempt to mitigate their feelings or expressions of disgust toward child molesters, their anger at terrorists, or their distrust of particular politicians. Quite opposite to the controlled processes of inhibition, people are likely to desire to “upregulate” their emotional responses to feel more negative (or less positive) if they have the opportunity. Although these groups have received less attention, there is evidence that these normatively stigmatized groups (e.g., obese people) also lead to activity in brain regions associated with affective (amygdala and insula) and controlled processing (ACC and lateral PFC; Krendl, Macrae, Kelley, & Heatherton, 2006). Whereas Krendl and colleagues found greater amygdala and insula—a region linked to the intense feelings, including disgust (Phillips et al., 1997)—activation to stigmatized groups compared with controls, they reported that there was *also* greater activity to these groups in the lateral PFC. Although the positive relationship between affective and controlled brain regions was interpreted as a failed attempt to control negativity, work by Ochsner and colleagues (2004) on the upregulation of emotion raises the possibility that lateral PFC activity in this study reflected an effort to increase or maintain a negative response that is personally or culturally acceptable.

Problems with Suppression and Alternate Approaches to Regulation

For the most part, investigations of prejudice regulation have focused on the ways that people can suppress automatically activated affective and stereotypical responses. Yet, in his analysis of emotion regulation strategies, Gross (Gross, 1998; Gross & Thompson, 2007) provided a useful taxonomy in which response-focused emotion regulation strategies are contrasted with antecedent emotion regulation strategies. Whereas response-focused strategies typically involve the simple suppression of an affective response following its activation, antecedent focused strategies attempt to shape an affective response prior to activation, or quickly following activation through processes of reappraisal. Interestingly, in comparing the pros and cons of each class of regulation, Gross notes that whereas antecedent forms tend to provide strong and adaptive changes in affective experience, response-focused strategies (and suppression in particular) tend to work only for short periods of time, are associated with unhealthy physiological side effects (such as high blood pressure) and, most important, can backfire and result in rebound effects.

Thus, an unfortunate consequence of a desire to suppress all prejudiced thoughts and feelings is that, to the extent that automatic bias is pervasive, people will need to engage in the most effort to control these biases and may therefore be the ones who suffer the largest cognitive costs (Baumeister, Bratslavsky, Muraven, & Tice, 1998). Providing evidence for this hypothesis, Richeson and Shelton (2003) found that after White participants with high levels of automatic racial bias interacted with a Black individual, they subsequently performed worse on the Stroop task, which requires cognitive control. Presumably, White participants with racial bias on implicit measures had the most bias to control, and therefore were cognitively depleted following an interracial interaction. These studies suggest that people with automatic racial biases need to engage in greater levels of controlled processing to successfully navigate interracial interactions and these extra efforts lead to subsequent impairments in controlled processing, raising doubts about their ability to suppress bias for any extended period. Ironically, people who try the hardest to suppress their biases may be the most likely to express these biases in later intergroup encounters.

In a follow-up fMRI study, Richeson and colleagues (2003) scanned White participants while they viewed Black and White faces during fMRI. Afterward, participants interacted with a Black confederate and then performed the Stroop task. As would be expected if participants were attempting to control prejudice while in the scanner, heightened activation to Black than White faces was observed in areas of right lateral PFC and ACC. More importantly, these levels of activation correlated with poorer performance on an executive function task (the Stroop task) following scanning. These patterns of results provide support both for the idea that nonprejudiced participants attempt to regulate their emotional responses to Black faces and that this regulation depletes executive functioning resources. As such, this provides strong evidence that the attempt to suppress by those most wanting to think and feel without prejudice may ironically be the ones who may be most likely to fall victim to automatic bias after their cognitive resources have been depleted.

Because suppression can lead to negative consequences, both for the social perceiver and the targets of prejudice, research is needed to examine alternative strategies for regulation that have the potential to avoid the unintended negative side effects of suppression. One unstudied but potentially promising approach is conscious reappraisal—the process of consciously changing the meaning and the appraisal of social groups and their members. Work by Kevin Ochsner and colleagues have consistently found that changing cognitions about an event or person changes affective responses (see Ochsner & Gross, 2005 for review). If people change their cognitions to feel more negative, greater amygdala activation is found, and if people change their cognitions to feel less negative, less amygdala activation is found (Ochsner, Bunge, Gross, & Gabrieli, 2002; Ochsner et al., 2004). This suggests that people can shape the contents of their mental space by foregrounding some pieces of information and backgrounding others to generate an emotional response that is consistent with their goals and values.

Another strategy for changing the way that evaluations unfold is for people to consciously change their processing goals (see Cunningham, Van Bavel, & Johnsen, 2008). A processing goal that may be particularly useful is an explicit motivation to individuate people and place less emphasis on group membership in person perception. In this context, one hypothesis is that having the goal to look for individuating characteristics may change the level of processing and may indirectly reduce the power of automatic stereotypes and prejudices. To test this prediction, Wheeler and Fiske (2005) presented White participants with Black and White faces and had them engage in judgments that were designed to induce participants to either process the faces as individuals or as members of social groups.

Consistent with previous research, when participants engaged in social categorization (e.g., classifying the faces by age), they showed greater amygdala activity to the Black than White faces. However, when participants were simply asked to decide whether each person preferred particular vegetables (a task thought to increase attention to individuating features), the amygdala no longer activated more to Black than White faces. To the extent that amygdala activation can be taken as affective bias, the simple act of individuation eliminated the standard race bias effect.

Summary

The study of intergroup relations using neuroscience methods remains relatively young. Nevertheless the past decade of research has revealed several important insights into the complexity of intergroup perception and evaluation. This research has provided exciting evidence of the automaticity of intergroup perception and evaluation, and the complex interactions between the component processes that guide behavior. These studies highlight the speed with which individuals distinguish different groups, their ability to do so without conscious awareness, and their ability to alter these initial processes according to motivations or goals. Improvements in technology and convergence across methods will add precision and contribute novel insights about an evaluative system that influences intergroup relations. In addition, the insights gleaned from a multilevel approach will eventually lead to novel predictions for traditional behavioral investigations and ultimately interventions that improve intergroup relations.

- 1 Not all people are motivated to control their prejudice to the same degree. Some people may strive to be egalitarian in all their thoughts and feelings, remaining vigilant at all times, whereas others may not care about their prejudiced reactions until they need to conceal it from others. Recent research explored the effect of internal versus external motivations to respond without prejudice on different aspects of controlled processing during the shooter task (Amodio & Devine, 2006). Replicating the previous research by Amodio and colleagues (Amodio et al., 2004), the more automatic dorsal ACC was linked to control of racial bias on the shooter task across conditions among White participants. However, the later rostral ACC component was only associated with control in the shooter task among participants with a high external motivation when they were in a public situation that precluded racially biased responding. That is, people motivated for social reasons only engaged in controlled processing mediated by the rostral ACC when they were in a situation where social constraints were a factor, and this aspect of control took slightly longer. These data suggest that more automatic aspects of control (dACC) were rapidly engaged and insensitive to contextual pressures, whereas external motivations and contextual pressures triggered more delayed aspects of controlled processing (rACC).

References

- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, 91, 652–661.
- Amodio, D. M., 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.
- 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., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and 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.
- Bernstein, M., Young, S., & 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.
- 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. (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 Behavior*, 9, 415–424.
- Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the “other-race” face classification advantage. *NeuroReport*, 15, 905–910.
- Cohen, J. D., Botvinick, M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: Who’s in control? *Nature Neuroscience*, 3, 421–423.
- Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., & Leyens, J.-P. (2005). Infrahumanization or familiarity?: Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin*, 31, 243–253.
- Crandall, C. S., Eshleman, A., & O’Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82, 359–378.
- Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, 30, 1332–1346.

- 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.
- 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 and Social Psychology, 82*, 62–68.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: 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 and Social Psychology, 50*, 229–238.
- 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. (in press). Understanding and using the Implicit Association Test: Pt. III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*.
- 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.
- Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science*, *13*, 135–141.
- Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*, 847–853.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup versus ingroup face stimuli. *NeuroReport*, *11*, 2351–2355.
- 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.
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*, 785–794.
- 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.
- 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.
- Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. D. Williams, & W. von Hippel (Eds.), *Social judgments: Explicit and implicit processes* (pp. 44–67). New York: Cambridge University Press.
- 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.
- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000, June 9). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.
- Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other “race.” *Journal of Personality and Social Psychology*, *13*, 330–334.
- Mitchell, J. P. (2006). Mentalizing and marr: An information processing approach to the study of social cognition. *Brain Research*, *1079*, 66–75.

- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50, 655–663.
- Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996, October 31). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, 383, 812–815.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dynamics*, 6(1), 101–115.
- Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, 14, 1215–1299.
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9, 242–249.
- Ochsner, K. N., Ray, R. D., Robertson, E. R., Cooper, J. C., Chopra, S., Gabrieli, J. D. E., et al. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, 23, 483–499.
- Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, 112, 536–552.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192.
- Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: 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., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia*, 41, 203–208.
- Phelps, E. A., 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.
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997, October 2). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495–498.
- Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience*, 16, 189–203.

- 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.
- Seeleman, V. (1940). The influence of attitude upon the remembering of pictorial material. *Archives of Psychology*, 36, 6–69.
- Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, 7, 36–97.
- Vaes, J., Paladino, M. P., Castelli, L., Leyens, J.-P., & Giovanazzi, A. (2003). On the behavioral consequences of infrahumanization: The implicit role of uniquely human emotions in intergroup relations. *Journal of Personality and Social Psychology*, 85, 1016–1034.
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 44A, 671–703.
- 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 .
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, 9, 585–594.
- 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.