CHAPTER FOUR

The neural bases of attitudes, evaluation, and behavior change

Emily B. Falk and Matthew D. Lieberman

INTRODUCTION

Attitudes encompass long-standing evaluations of people, places, and ideas, and may influence a range of behaviors, including those that directly impact political behavior, intergroup relations, and health behaviors, among other consequences. Attitudes are central in answering questions such as: Where should we invest community resources? Whom should we vote for in the next election? Where will we spend our paychecks? As such, the study of attitudes has captivated thinkers for centuries, and scientists for decades (Allport, 1935; Aristotle, 1924/1954; Hovland, 1949; Hovland, Janis, & Kelley, 1953). Gordon Allport (1935) called attitudes "the most distinctive and indispensable concept in contemporary American social psychology" (p. 798), and suggested that understanding attitudes would allow us to understand not only the preferences and behaviors of individuals, but would also provide broader insight into the actions of groups and cultures. With this in mind, Allport (1935) defined an attitude as "a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon an individual's response to all objects and situations with which it is related" (p. 810).

Following this early work, research has continued to build our understanding of attitudes and attitude change (Albarracin, Johnson, & Zanna, 2005; Eagly & Chaiken, 1993, 2005; Petty & Cacioppo, 1986; Petty, Priester, & Wegener, 1994), however, many questions concerning the nature of attitudes, as well as the underlying mechanisms of attitude formation and attitude change remain unanswered (Eagly & Chaiken, 2005; Gawronski, 2007). For example, what is the role of implicit attitudes in influencing explicit attitudes, behaviors, interactions with other individuals and groups? How do people internally regulate unpleasant or undesirable attitudes and biases? What are the mechanisms

through which attitudes form and change, and what are the mechanisms through which external influences, such as persuasive appeals, influence attitudes and behaviors? Although these are clearly complex questions, they are made even more challenging to tackle by demand characteristics, participants' self-presentational concerns and the fact that participants may not have conscious awareness of the ways in which they are processing information. Together, all of these factors may lead to biased self-reports (Krosnick, Judd, & Wittenbrink, 2005).

As foreshadowed by Allport's (1935) definition of attitudes, which includes a "neural state of readiness, organized through experience," the brain may be able to shed some light on unanswered questions that introspection and self-report data have not (Lieberman, 2007, 2010; Ochsner & Lieberman, 2001). More specifically, the recent advance of neuroimaging technologies has opened new possibilities to examine multiple psychological processes in concert, to examine the extent to which different phenomena share common or distinct neural bases, and to link theory developed in social psychology to an extensive neuroscience literature developed in human and animal models.

For example, a vast literature on fear, conditioning, and social behavior in animals has been key in informing existing theories of prejudice, bias, and social behavior in humans (Amodio & Lieberman, 2009). Furthermore, our evolving understanding of the neural bases of automatic and controlled processes has provided insight into the ways in which implicit and explicit evaluations and attitudes interact. A body of literature is also beginning to form examining the neural correlates of closely related concepts such as the subjective experience of persuasion, attitude change, behavior change, and message propagation. Lastly, the literature addressing the neural mechanisms that support attitudinally relevant processes has reached a stage where integration can begin to take place (Cunningham & Zelazo, 2007; Cunningham et al., 2007).

Prominent theorists since Allport have also worked from a relatively broad definition of attitudes as evaluative tendencies that can have cognitive, affective, and behavioral antecedents and consequences (Eagly & Chaiken, 2007); in this chapter we will explore the ways in which neuroscience informs our understanding of these processes. This chapter is divided into three main sections: The Neural Bases of Responses to Outgroups and the Regulation of Bias; The Neural Bases of Evaluation and Preferences; and The Neural Bases of Persuasion, Attitude, and Behavior Change.

THE NEURAL BASES OF RESPONSES TO OUTGROUPS AND THE REGULATION OF BIAS

Much early neuroimaging work exploring the neural bases of attitudes was in the context of race-related attitudes and intergroup relations. In many ways,

race-related attitudes are similar to other types of attitudes. For example, they may have affective, cognitive and behavioral components, and can be subject to conscious reflection or may reside under the surface. However, strong societal norms surrounding race and prejudice as well as self-presentation concerns on the part of participants create methodological challenges in determining peoples' "real" racial attitudes.

Proxy measures of implicit attitudes such as the implicit association test (IAT) have thus far been one of the few means of inferring what individuals are unable, or unwilling to self-report. The IAT measures strength of association between concepts through a process of timed categorization; objects and evaluative words are typically paired (e.g., in a first round, a left button might be simultaneously associated with words that are "good" and faces that are black, and the right button with words that are "bad" and faces that are white, whereas in a second round, the pairings would be reversed). It is thought that category pairings that are more strongly associated in memory will result in faster reaction times (Bargh et al., 1992; Draine & Greenwald, 1998; Fazio et al., 1986; Greenwald & Banaji, 1995). Like all measures, however, the IAT has limitations (Karpinski & Hilton, 2001; Rothermund & Wentura, 2001, 2004), and scientists have sought complementary methods for investigating implicit and automatic processes. Researchers commonly observe discrepancies between implicit and explicit attitudes surrounding race, and between self-reported attitudes and observed behaviors.

Most often, individuals simultaneously report unbiased attitudes, but behave in biased ways. Indeed, old-fashioned racism has decreased in the United States since Allport's time. However, a majority of white Americans still exhibit a preference for whites over blacks on implicit evaluation measures (Chen & Bargh, 1997; Devine, 1989; Nosek, Banaji, & Greenwald, 2002); even individuals who hold explicitly non-racist attitudes and believe in equality may demonstrate biased behaviors towards outgroup members (Amodio et al., 2006; Dovidio, Kawakami, & Gaertner, 2002). Many of these behaviors are linked to implicit attitudes, with implicit and explicit attitudes predicting different types of biased behavior, and with implicit racial categorization taking place even when race is irrelevant to task demands (Dickter & Bartholow, 2007; Fazio et al., 1995).

There are several possible explanations for the discrepancy between implicit and explicit attitudes, and explicit attitudes and behaviors (for a review, see Amodio & Lieberman, 2009). One is that Americans are still just as prejudiced, but that social norms now preclude the outward expression of racism. A second possible explanation is that even participants who do not hold conscious prejudice have learned cultural associations with different racial groups (e.g., blackness and whiteness). Knowledge of cultural stereotypes may be reflected in response to implicit tasks (e.g., reaction time tasks such as the IAT), which by definition tap into our fast, automatic associations, as well as in more subtle

behaviors that are outside of conscious control (e.g., body language). To this point, several research teams have harnessed neuroimaging as a method for exploring responses to racial outgroups, focusing heavily on the amygdala as a key correlate of race bias. This stems from the amygdala's role in fear conditioning (Davis, 1992), and the hypothesized relationship between fear, threat, and prejudice (Smith, 1993). It should be noted that these findings often rely on reverse inference, and, as such, should be interpreted cautiously (Poldrack, 2006); the presence of an automatic limbic response does not necessarily reflect prejudice or fear, and indeed, recent findings suggest that the amygdala and other limbic structures may reflect motivational relevance more broadly (Van Bavel, Packer, & Cunningham, 2008).

The first study to explore the relationship between implicit and explicit racial attitudes in the brain was conducted by Phelps and colleagues (Phelps et al., 2000). In this study, white participants viewed photos of black and white male faces as part of a task that was unrelated to social evaluation. The researchers then had participants complete both an explicit measure of modern racism (the Modern Racism Scale, McConahay, 1986), and two implicit measures of race bias (the IAT and a startle eye blink task). Whereas most participants did not show any bias on the explicit racism measure, many did show bias according to the implicit measures. Interestingly, there was no main effect of black versus white faces on brain activity across participants, however, the amount of bias expressed through implicit measures was positively correlated with amygdala activity.

Subsequent research has also demonstrated relationships between implicit bias and amygdala activity for both racial and non-racial (e.g., political) outgroups (Amodio et al., 2004; Cunningham et al., 2004a; Eberhardt, 2005; Hart et al., 2000; Kaplan, Freedman, & Iacoboni, 2007; Phelps, 2001; Phelps, Cannistraci, & Cunningham, 2003; Phelps & Thomas, 2003). This work is typically characterized as exploring responses to outgroups (and is referred to as such in this chapter), however, most of the studies reviewed examine responses of white participants to black faces.

Breaking this pattern, work by Lieberman and colleagues (Lieberman et al., 2005) using both black and white participants demonstrated that both black and white participants showed increased amygdala activity in response to black faces, suggesting that cultural learning, and not the familiarity of one's own race, may be responsible for the responses observed. This is also consistent with behavioral work demonstrating that black participants often hold implicit biases against black targets (Ashburn-Nardo, Knowles, & Monteith, 2003; Livingston, 2002), and with neuroimaging findings that increased amygdala activity is observed when white people respond to photographs of darker skinned white people as compared to lighter skinned photographs of white people (Ronquillo et al., 2007). By contrast, in a recent study (Dickter & Bartholow, 2007) examining attention biases to race targets revealed by event related potentials, both

main effects of the race of target faces, as well as target race x participant race effects were demonstrated, reinforcing the importance of accounting for both target race and participant race. Also consistent with this view, implicit bias and corresponding neural responses are not constant across situations (Cunningham et al., 2004a; Lieberman et al., 2005). In the next section, we turn our attention to address when and how individuals are likely to regulate automatic, biased responses.

Regulation of race bias

A number of questions inform our understanding of when and how individuals regulate automatic, biased responses. As examples, researchers have considered questions such as: Is fear/arousal an uncontrollable response to outgroup members? If not, how do individuals regulate these responses? In cases when individuals do not exhibit biased behaviors, is it the case that an automatic biased response has been successfully regulated or can bias be prevented before it begins? Under what circumstances do automatic biases predominate, and under what circumstances should we observe more controlled processing? Finally, how are different regulation strategies related to different behavioral outcomes?

Neural measures have proven useful in characterizing initial responses to outgroups, as well as regulatory processes that follow. For example, Phelps and colleagues not only examined amygdala activity to outgroup faces, but also whether people spontaneously regulate this response under some circumstances. Their data showed that whereas unfamiliar black faces elicited more amygdala activity than unfamiliar white faces, this effect disappeared when both white and black faces were positively perceived, familiar faces (Phelps et al., 2000). Similarly, Wheeler and Fiske (2005) observed that white participants showed increased amygdala activity in response to black versus white faces when asked to categorize the race of the person presented (race salient condition), but this effect disappeared when participants were asked to personalize the individuals depicted by guessing information about the target, such as whether the target liked various vegetables (Wheeler & Fiske, 2005). Likewise, in a recent study participants made superficial ratings (regarding age) or personal ratings (regarding food preferences) of stigmatized outgroup members. When making superficial judgments, increased activity was observed in affective processing regions (e.g., insula), whereas increased activity in social-cognitive/self-processing regions (e.g., medial prefrontal cortex, MPFC) was associated with making individuating, personal ratings (Harris & Fiske, 2007). Results of this kind are consistent with the idea that individuation of outgroup targets may reduce the automatic tendency toward bias.

A study by Cunningham and colleagues examined the conditions under which intentional regulation of bias is likely to occur (Cunningham et al.,

2004a). In this study, white participants viewed photos of black and white human faces for short (30 ms) or longer (525 ms) time periods while in a functional Magnetic Resonance Imaging (fMRI) scanner. The shorter duration stimuli were not accessible to conscious awareness (participants did not report seeing them). When participants viewed black faces (compared to white faces), participants showed increased amygdala activity in response to black faces when the stimuli were presented outside of conscious awareness. However, when the participants had the opportunity to consciously process the stimuli (when the face was on the screen for 525 ms) the difference in the amount of amygdala activity to black versus white faces was reduced, and activity in areas of controlled processing (right ventrolateral PFC, VLPFC; right dorsolateral PFC, DLPFC; anterior cingulate cortex, ACC) increased. Furthermore, activity in controlled processing regions such as the DLPFC and ACC was inversely correlated with change in amygdala activity, suggesting that these areas may be recruited to downregulate the initial amygdala response. Therefore, the authors suggest that when viewing members of an outgroup, initial responses tend to be automatic and affective, but that this response is soon regulated by more controlled processing in the PFC and ACC (Cunningham et al., 2004a). Given that all participants reported low levels of prejudice on an explicit measure, it is likely that participants were motivated to present themselves as non-prejudiced,[1] and quickly regulate the initial automatic, affective response. This pattern of results has also been observed in response to other stigmatized outgroups (e.g., obese, transsexual, unattractive, and facially pierced individuals), with increases in affective processing regions such as the amygdala and insula prompting greater responses in regulatory regions such as the ACC and PFC (Krendl et al., 2006). Over time, however, affective responses and the regulation thereof may be subject to familiarity as well; Hart and colleagues observed that amygdala activity was initially similar in response to both unfamiliar ingroup and outgroup faces, but habituated more quickly to ingroup faces (Hart et al., 2000).

Other deliberate factors can influence the use of controlled processing to regulate bias as well. For example, Lieberman and colleagues required participants to either match images on the basis of race or label the race of faces presented (Lieberman et al., 2005). The authors reasoned that the top-down nature of the verbal labeling task would require more controlled processing, and indeed in this study, the verbal labeling task showed increased activity in right VLPFC, a neural region often implicated in emotion regulation. Furthermore, although the authors observed increased amygdala activity when participants visually matched photos of people according to race, the effect disappeared when participants were required to verbally label the images as belonging to a given race, and the amount of increased activity in right VLPFC correlated with decreases in amygdala activity. The authors conclude that although automatic responses are likely when individuals are confronted with images, the process of labeling these evaluatively laden stimuli has a top-down regulatory effect (Lieberman et al., 2005).

Consequences of the need to regulate

The strength of the relationship between automatic neural responses in the amygdala and their regulation by prefrontal networks prompted Richeson and colleagues (2003) to hypothesize that for people who have a strong, automatic tendency toward implicitly biased attitudes, regulation might become more difficult under conditions of cognitive load or when controlled processing resources are otherwise depleted. They hypothesized that even participants who hold explicitly unbiased attitudes, and who are likely to be motivated to regulate that bias (due to societal norms, or for other reasons), might show increased bias following a demanding cognitive task. Likewise, following an interaction with an outgroup individual, participants might show evidence of depleted cognitive resources (operationalized by interference on an ostensibly unrelated Stroop color-naming task). Indeed, in a series of behavioral and fMRI studies, this is exactly what they found (Richeson et al., 2003; Shelton et al., 2005). In the fMRI portion of the investigation, the extent of controlled processing (as indexed by activity in DLPFC) engaged by the presentation of black faces was correlated with implicit racial bias. This suggests that for individuals who held greater implicit bias, more prefrontal resources were recruited when confronted with a situation that warranted regulation. Furthermore, the amount of activity in prefrontal cortex engaged by presentation of black faces mediated the relationship between implicit bias and interference on the cognitively demanding Stroop color-naming task following interaction with a black person, providing evidence for the hypothesized depletion mechanism. These results provide additional support for the idea that activity in regions that are typically associated with controlled processing can regulate automatic race bias. However, prolonged need to regulate in one area may spill over to produce decreased regulatory ability in other tasks (Richeson et al., 2003).

Finally, individual differences are also observed in the tendency to automatically regulate bias. In one study of low prejudice individuals (selected for high Internal Motivation to Respond Without Prejudice [Plant & Devine, 1998]), participants were led to believe either that their responses would remain confidential (private), or that the experimenter would monitor their responses to assess whether the participant appeared prejudiced (public). In this study, activity in neural regions linked to conflict detection (dorsal ACC) predicted stereotype inhibition in both private and public settings. However, activity in neural regions associated with error-perception (rostral ACC) predicted behavioral control of bias for individuals who reported high sensitivity to societal non-prejudice norms in public settings (Amodio et al., 2006).

Summary

Whereas early studies examining amygdala responses to black faces produced conflicting results, these discrepancies may be resolved by understanding the time course of the stimulus presented, and the sensitivity of the experimental design to pick up on changes in key brain regions over time, and individual differences in disposition or situational constraints. Greater demands on controlled processing resources may deplete the ability for participants to regulate bias. The amount of implicit bias observed is likely a function of interplay between the strength of automatic responses (indexed by activity in affective processing regions such as the amygdala), and the strength of controlled processing (indexed primarily by activity in networks involved in controlled processing in PFC). Processing of outgroups and other stigmatized categories is influenced both by the time course of the stimulus and response measured, by the demands of the task at hand (Cunningham et al., 2007), and by factors such as prior contact with the outgroup (Walker et al., 2008). Furthermore, specific regulation strategies can be employed to reduce implicit biases that would otherwise be present (Harris & Fiske, 2007; Krendl et al., 2006; Lieberman et al., 2005; Wheeler & Fiske, 2005).

Stereotypes, bias, and non-racial outgroups

Race is one of the most salient characteristics that distinguish people in groups. Therefore, much of the work relating neurocognitive activity to bias and the regulation of bias has focused on race (Lieberman, 2007). However, other work has explored the extent to which these same processes apply in other intergroup situations (e.g., political outgroups, gender outgroups, etc.).

One particular area of interest has been neural responses to political outgroups. Consistent with the literature on responses to racial outgroup faces, several researchers have examined the interplay between automatic, affective responses and more controlled, deliberate responses to partisan outgroup faces. For example, Knutson and colleagues observed that the activation of political attitudes (operationalized as performing an IAT using images of politicians who belonged to the participant's ingroup and outgroup) produced neural activations in both emotion processing regions and regions of lateral PFC implicated in deliberative reasoning. Participants who reported stronger party affiliation on an explicit measure of political preferences, however, showed less activity in controlled processing regions (lateral PFC) while completing the IAT. These findings are consistent with the idea that political attitudes may be processed along stereotypical or symbolic lines in cases when individuals hold stronger prior attitudes (Knutson et al., 2006). Recent work has also examined the ways in which humans process different forms of political beliefs (Zamboni et al., 2009). Zamboni and colleagues examined neural responses to political

statements that varied in the extent to which they emphasize: the good of an individual *vs.* the good of society (a dimension which the authors label "individualism"); liberal *vs.* conservative views (a dimension which the authors label "conservatism"); and moderate *vs.* radical solutions to government (a dimension which the authors label "radicalism"). The authors report that when reflecting on statements that are high on the individualism dimension, participants showed increased activity in a region associated with self-related processing (ventromedial PFC, VMPFC), whereas when reflecting on statements at the opposite end of the scale (high on value to society), participants evidenced increased activity in mentalizing regions including dorsomedial PFC (DMPFC) and temporoparietal junction (TPJ). In this study, processing more conservative statements was associated with increased activity in DLPFC, which the authors speculate may reflect increased cognitive processing related to self-interest/fairness dissonance or processing of other party views (given that most participants were liberal). Finally, the authors report that processing more radical statements was inversely associated with activity in the ventral striatum (VS). The authors note that VS is often implicated in reward processing (and speculate that inverse relationship between processing radical statements and VS activity may reflect the normative value of less radical beliefs). The authors also suggest that their findings are consistent with accounts of motivated reasoning in political information processing (Westen et al., 2006).

As in the literature describing race bias, however, some types of information processing are more likely to lead to automatic, affective responses than others. For example, consistent with the work of Lieberman et al. (2005), Knutson et al. (2006) observed increased amygdala activity when participants viewed images of outgroup politicians' faces during the IAT, but not when responding to written names. A study by Kaplan et al. (2007) also showed activity in both automatic and controlled processing regions in response to the presentation of outgroup political faces, but these authors came to different conclusions regarding the meaning of the activated networks. Activity in affective processing regions (e.g., insula), as well as more control oriented regions of PFC was observed when viewing the faces of political candidates from an opposing political party as compared to viewing faces of the participant's own political candidate. However, unlike previous work on race bias, in which prefrontal areas are interpreted as being engaged to downregulate negative affective responses, the authors of this study suggest that the presentation of political outgroup faces actually engages controlled processing networks that *upregulate* negative affective responses (Kaplan et al., 2007). Unfortunately, the temporal resolution of the study did not allow causal inference about the direction of the effect, but it is interesting to consider this interpretation in light of the differing motivations inherent to the stimuli; people are motivated to appear less racist, but it may be more societally acceptable to be a strong partisan.

In line with this hypothesis, research on motivated reasoning suggests that when reasoning about counter-attitudinal political information (e.g., strong partisans reasoning about information that is threatening to their preferred candidate), increased activity is observed in regions associated with affective evaluation (VMPFC, amygdala, insula), and regions associated with self-related processing (VMPFC, precuneus/posterior cingulate) but not cognitive control areas such as DLPFC and ACC (Westen et al., 2006). Bruneau and Saxe (2010) also report that precuneus activity is associated with bias in the context of Arabs and Israelis reading statements from ingroup and outgroup members. In this case, precuneus activity distinguished between ingroup and outgroup statements, and strength of activity in this region was associated with the degree of bias recorded on both explicit (feeling thermometer) and implicit (IAT) measures of outgroup bias (Bruneau & Saxe, 2010).

More broadly, elements of stereotyping and prejudice may be supported by the same neural mechanisms that support automatic evaluative processing more generally (e.g., VMPFC), an issue that will be discussed in more detail later in this chapter. For example, early work by Milne and Grafman (2001) explored gender stereotyping effects in patients with VMPFC damage. In this case, both VMPFC patients and healthy controls demonstrated equal gender stereotyping on an explicit measure; however, when performing a gender relevant IAT, patients with ventral prefrontal damage did not show the stereotypic gender associations displayed by healthy control subjects (Milne & Grafman, 2001). Subsequent follow-up work in lesion patients further distinguished between the medial and lateral aspects of ventral PFC (Gozzi et al., 2009); in this work, the degree of medial damage was correlated with increases in stereotypic implicit attitudes, whereas lateral damage was associated with decreased stereotypic implicit attitudes. This work also converges with fMRI evidence that activity in VMPFC (and amygdala, among other regions) is related to gender stereotyping (Quadflieg et al., 2009).

Thus, consistent with responses to racial outgroups, individuals also tend to demonstrate automatic and affective responses to non-racial outgroups, which are correlated with increased neural activity in areas such as the VMPFC, amygdala, and insula. Also consistent with the literature on race processing, in many cases, changes in affective processing regions are accompanied by corresponding changes in regulatory regions of lateral PFC. Unlike the responses to racial outgroups, however, there may be less societal pressure to appear unbiased toward non-racial outgroups (e.g., partisan groups), and hence different regulatory strategies may be employed.

Summary

Research regarding neural responses to outgroups initially focused on the amygdala as a key correlate of the threat response. Building on this work, other

affective processing regions, including insula and VMPFC have also been observed in response to group-based evaluative tasks. In contrast, areas of lateral PFC have been described as key correlates of the regulatory response. When considering responses to non-racial outgroups (e.g., in the context of political figures), research has also addressed the possibility that belief processing is not unidimensional, and people may be motivated to either down- or upregulate their automatic evaluative tendencies.

THE NEURAL BASES OF EVALUATION AND PREFERENCES

In recent years, social cognitive neuroscience has broadened the range of attitude objects studied using neuroimaging technology, and as such has broadened our understanding of the processes that lead to evaluation and preference more generally. Evaluation and preference are central in defining the nature of attitudes (Eagly & Chaiken, 1993, 2005, 2007; Petty et al., 1994; Zajonc & Markus, 1982), although scholars disagree about the extent to which these evaluations and preferences must remain stable to be considered "real" (Bishop, 1980; Bishop, Hamilton, & McConahay, 1980; Converse, 1970). Regardless of the definition used, however, the relationship between evaluations, preferences, and attitudes is intertwined; underlying attitudes may predispose individuals to evaluate objects, situations, people, or groups more or less favorably, and depending on the evaluation that is made, individuals may update their underlying attitudes.

However, this process cannot be directly observed. Put another way, current behavioral researchers "do not have an inherent psychological reality that can be verified. In other words, researchers cannot directly observe object-evaluation associations, knowledge structures, or microconcepts" (Eagly & Chaiken, 2005, p. 746). This is especially true in the case of implicit and unconscious attitudes. Whereas people sometimes deliberately evaluate the world around them, they also automatically and spontaneously make evaluations that are outside the realm of awareness. Thus, the resulting attitudes may differ in important ways from consciously and deliberately formed opinions. Several studies exploring the neural basis of evaluative judgments and preferences focus on explaining differences between the processing of implicit and explicit judgments and the expression of implicit and explicit attitudes and preferences.

Automatic and controlled processing in evaluation and preference

Across a range of domains including judgments of beauty (Jacobsen et al., 2005; Vartanian & Goel, 2004), evaluations of places, events, and political figures (Zysset et al., 2002), and brand preferences (McClure et al., 2004), when people

make explicit evaluations, regions typically associated with controlled processing, including areas of MPFC, VLPFC, medial parietal cortex (MPAC), lateral parietal cortex (LPAC), and ACC are engaged (Lieberman, 2010). By contrast, when tasks do not require explicit evaluative judgments, increased activity is observed in regions typically associated with automatic processing, such as the amygdala and VMPFC, as well as the insula (Lieberman, 2010). This is true for tasks that involve targets that are presented subliminally and when preferences are measured through implicit behavioral means.

Supporting the role of the VMPFC in implicit evaluation, Koenigs and Tranel (2008) showed that when asked to perform a blind taste test (without brand labels) of Coke versus Pepsi, patients with damage in the VMPFC, healthy controls, and patients with non-VMPFC brain damage showed a preference for Pepsi. However, in an open taste test with brand labels, healthy controls and patients with lateral brain lesions show a preference for Coke (the so-called "Pepsi paradox"), whereas patients with VMPFC damage maintained their original choices, failing to show typical brand preference effects (Koenigs & Tranel, 2008). Thus, patients with VMPFC damage did not show the characteristic "Pepsi paradox" effect, suggesting that VMPFC may be partly responsible for "translating commercial images into brand preferences" (Koenigs & Tranel, 2008, p. 1).

In a study of implicit and explicit evaluations of people, Cunningham and colleagues asked participants to explicitly evaluate famous people on a good/bad dimension (e.g., Hitler = bad), while on other separate trials participants classified famous people as past/present (e.g., Hitler = past) (Cunningham et al., 2003). When comparing brain activity associated with explicit evaluation (Hitler = bad) in contrast to past/present classification (Hitler = past), the researchers observed increased activity in controlled processing regions such as MPFC and VLPFC, as well as the ACC. However, regardless of the intention to evaluate (in both the explicit good/bad evaluative and past/present classification conditions), increased amygdala and insula activity was observed in response to images of famous people who were considered "bad" compared to individuals who were considered "good," suggesting the likelihood of negative, affective processing (Cunningham et al., 2013).

Likewise, in a study in which participants evaluated a series of concepts (e.g., murder, happiness, abortion, welfare) on a good/bad dimension as well as categorized concepts on an abstract/concrete dimension, amygdala activity was positively correlated with the emotional intensity of the stimulus, and insula activity correlated with affective valence across conditions, regardless of intention to evaluate. However, when explicitly evaluating the ideas on a good/bad dimension, to the extent that participants said that they felt ambivalent or reported that they tried to control their evaluation of the topic, increased activity was observed in regions implicated in controlled processing such as the ACC, frontal poles, and VLPFC (Cunningham, Raye, & Johnson, 2004b).

These results support the idea that regardless of whether an explicit judgment is made, the brain processes the valence of stimuli, but depending on the demands of a given task or situation, conflicting information and explicit judgments recruit higher-level brain networks that may be more sensitive to attitudinal complexity.

Integration of automatic and controlled processing in evaluation and preference

Neuroimaging research has allowed scientists to simultaneously explore automatic and controlled processing, and hence to dissociate circumstances under which each is likely to predominate. It is not necessarily the case, however, that automatic and controlled aspects of evaluation are unrelated. In fact, it is likely that evaluations and expressed attitudes at any given time point are the product of interactions between neurocognitive networks that support automatic processing and networks that support controlled processing. In an effort to integrate the information brought to the fore by neuroimaging regarding the specific ways in which people make evaluations, form attitudes, and change those attitudes, Cunningham and Zelazo and colleagues have proposed an "iterative reprocessing model" of information and affective processing. The Iterative Reprocessing Model posits that at any given time point, evaluations are constructed from an interaction of faster automatic processes (subserved by limbic structures such as the amygdala), and controlled processes (subserved by structures in the PFC and parietal cortex). Cunningham and Zelazo (2007) propose that the way in which these two systems come together depends on time constraints, motivations, and situational factors, and that information is iteratively processed and reprocessed to arrive at an evaluation at any given point (Cunningham & Zelazo, 2007; Cunningham et al., 2007). The authors suggest that we need not conclude (as past research in social cognition has) that implicit and explicit attitudes are fundamentally different entities, but instead that automatic evaluations are important across iterations and are influenced by, as well as influence, more controlled processes.

A recent study by Tusche and colleagues also provides support for overlap in the systems that encode explicit and implicit attitudinal information. In their work, one group of participants was presented with images of cars and explicitly asked to attend and evaluate the cars. A second group of participants also viewed images of the same cars, while performing a distracter task, with no explicit instructions to attend to, or evaluate, the cars. Following the scanner session, both groups of participants were asked to picture themselves in a situation in which they might purchase a car, and provided ratings of how likely they would be to purchase each car. In this study, neural activity in VMPFC and insula were associated with later consumer choice ratings, in both the explicit attention and implicit attention group (Tusche, Bode, & Haynes, 2010).

Summary

Thus, regardless of intention to evaluate, the brain seems to register an affective (potentially evaluative) response to target objects in areas such as the amygdala and insula, and the VMPFC appears to integrate value signals. Under circumstances in which a more controlled reaction or explicit evaluation is required, areas of the brain that tend to be more involved in controlled processing and conflict monitoring, such as the lateral PFC, parietal cortex and ACC, become involved. Researchers have suggested that the two systems interact over a series of iterations in the brain, and that final evaluations are a function of factors such as time constraints, motivations, and the external situation.

THE NEURAL BASES OF PERSUASION, ATTITUDE, AND BEHAVIOR CHANGE

Having considered the ways that the brain supports our evaluations of objects, concepts, brands, people, and groups, in this final section we consider the ways in which neural activity informs our understanding of the consequences of implicit and explicit evaluation. More specifically, we will explore the neuroscience of attitude and behavior change. We will briefly explore both an example of internally driven attitude change (cognitive dissonance), as well as external factors that influence behavior (the subjective experience of persuasion, and neural predictors of behavior change in response to persuasive messages).

Dissonance based attitude change

Early work in social cognitive neuroscience to explore attitude change was conducted by Lieberman and colleagues (Lieberman et al., 2001). This work explored the phenomenon of cognitive dissonance, in which conflicting initial attitudes and behaviors are believed to produce discomfort that leads to subsequent attitude change (Festinger, 1957). This work examined dissonance induced attitude change in both anterograde amnesia patients and healthy controls. Whereas the amnesia patients had no memory of performing a behavior that conflicted with their prior attitudes, the patients changed their attitudes to be more in line with the performed behavior just as healthy controls did. Thus, in contrast to previous explanations of cognitive dissonance effects involving conscious rationalization, the researchers suggested that even when individuals have no memory of inconsistent prior attitudes and behaviors, implicit processes are likely at work that still result in attitude change.

Subsequent imaging work has also explored post-decisional attitude change (dissonance), wherein two similarly valued alternatives are presented and

participants are forced to choose between them. In this context, after making a choice, the chosen object is subsequently valued more highly than the unchosen object. Consistent with the work of Lieberman and colleagues (2001), activity in automatic, affective processing regions (e.g., striatum) prior to the choice predicts which alternative is likely to be chosen (Sharot, De Martino, & Dolan, 2009), even though this information is not accessible to conscious awareness. Furthermore, post-decisional reward processing is even greater in response to the chosen versus unchosen alternatives, suggesting that the neural response is altered by the degree of commitment to the attitude object (Sharot et al., 2009). In parallel, work by Jarcho and colleagues suggests that neural activity associated with self-control (right VLPFC) and subjective valuation (VMPFC, VS) is correlated with increased post-decisional, dissonance induced attitude change (Jarcho, Berkman, & Lieberman, 2011). Finally, a third group exploring the neural bases of dissonance effects (van Veen et al., 2009) reported that neural activity in the anterior insula and dorsal ACC predicts post-dissonance attitude change. Broadly speaking, each of these reports fit within the framework of attitude (or bias) regulation discussed in previous sections; initial automatic responses in affective processing regions are altered following a deliberate choice. The latter two research teams each suggest that neural circuitry involved in controlled processing may serve this regulatory role, altering the effect observed in affective processing regions, and potentially resulting in the observed dissonance effect. Future work will determine the circumstances under which each specific type of processing is likely to occur.

Persuasion

In considering phenomena such as evaluations of outgroup faces, partisan group symbols, and post-decisional attitude change, we have largely ignored the potential influence of outside sources intended to shape or change people's attitudes. However, many factors including societal norms, group norms, and explicit persuasive appeals influence individuals' attitudes and behaviors.

Preliminary work has begun to uncover the neural bases of the experience of persuasion by an external source (Chua et al., 2009a, 2009b; Falk et al., 2010b; Klucharev, Smidts, & Fernandez, 2008). Falk and colleagues demonstrated that across two diverse cultural/linguistic groups (Americans and Koreans) and using two different types of media (plain text and video-based messages), activity in the DMPFC, bilateral posterior superior temporal sulcus (pSTS), and bilateral temporal poles (TP) is associated with the experience of persuasion (Falk et al., 2010b). Furthermore, in some situations, medial temporal lobes, left VLPFC, VMPFC and visual cortex were correlated with the experience of persuasion. Likewise, Klucharev and colleagues observed that expert power in presenting arguments resulted in increased activity in left prefrontal and parietal cortices, as well as the medial temporal lobes, which

they attribute to increased semantic processing and memory encoding when information comes from an expert source (Klucharev et al., 2008). Finally, work by Chua and colleagues demonstrated that personalized messages elicited more activity in self-related processing regions such as MPFC and precuneus (Chua et al., 2009a, 2009b), messages with high information value elicited more activity in lateral prefrontal regions involved in reasoning, and motivational messages elicited increased activity in VMPFC, a region discussed earlier in this chapter to be involved in implicit valuation and affective processing (Chua et al., 2009b).

Interestingly, the constellation of regions observed most consistently in the persuasion studies conducted by Falk and colleagues (DMPFC, pSTS, TP), has previously been observed in response to tasks related to theory of mind processing (Frith & Frith, 2003), and not co-activated in response to other types of tasks (Cabeza & Nyberg, 2000). Regions involved in social cognition have also been associated with message propagation following exposure to persuasive messages (Falk et al., under review). Prior behavioral research has also touched on the relationship between perspective taking and persuasion (Aaker & Williams, 1998; Campbell & Babrow, 2004), but there has been less direct emphasis on social cognition as a key factor in persuasion research. Combined with the work of Chua and colleagues, further exploration of specific self-processes and social processes in the process of persuasion and message propagation may lead to fruitful results.

Behavior change

A final area of interest in considering how the brain supports evaluation, preference, and attitude change is whether the neural regions associated with making evaluations and changing attitudes also map onto the areas of the brain that predict relevant behavior changes. A number of studies suggest that VMPFC may play a key role in integrating value signals; activity in VMPFC and VS have been associated with predicting a number of proximal attitude and behavior outcomes including purchase decisions and willingness to pay (Knutson et al., 2007; Plassmann, O'Doherty, & Rangel, 2007). VMPFC also appears to track social information about attitude objects, which is used in making such evaluations (Mason, Dyer, & Norton, 2009; Plassmann et al., 2008), and integrates this information with other information sources (Hare, Camerer, & Rangel, 2009).

Extending this work to predict behavior outside of the fMRI environment, Falk and colleagues investigated whether neural activity in MPFC/VMPFC could predict behavior change over a longer time course, following exposure to persuasive messages. Participants' sunscreen use was recorded for a week prior to and following an fMRI scan in which participants were exposed to public service announcements about the need to wear sunscreen on a daily basis.

In this study, neural activity in MPFC/BA10 during message exposure was predictive of changes in sunscreen use from the week prior to the week following the scan (Falk et al., 2010a). Furthermore, neural activity explained approximately 23 percent of the variability in behavior change, above and beyond self-reported attitudes and intentions. This finding suggests that, as in studies of implicit preferences and attitude change, neural activity in VMPFC may index attitudinal or intentional precursors of behavior change that are outside of conscious awareness.

In an effort to explore whether neural activity in MPFC/VMPFC would also predict behavior change in the context of more complex and motivationally relevant behavior, Falk and colleagues (2011) monitored neural activity during exposure to advertisements designed to help smokers quit smoking, in a group of smokers who were committed to quitting. In this study, neural activity again explained considerable variance in behavior change (~20 percent), above and beyond self-reported intentions, self-efficacy, and ability to relate to the ads (Falk et al., 2011). This provides further evidence that the ventral portion of medial BA10/MPFC/VMPFC contains information about processes that may be inaccessible to self-report. Falk and colleagues (2012) also found that neural activity in this region predicted the success of messages at the population level, better than the self-reported projections of focus groups, and the projected efficacy assessed by experts in the field. Thus, it is possible that neural activity may not only reveal information about likely behavior change in individuals whose neural activity is being recorded, but that this information also extends to larger groups of people at the population level (Berns & Moore, 2012; Falk, Berkman, & Lieberman, 2012).

Summary

VMPFC appears to play a key role in implicit valuation of stimuli, and in integrating value signals along different dimensions. Activity in VMPFC predicts proximal outcomes such as willingness to pay, as well as longer-term outcomes including health behavior change. It also appears that signals in VMPFC in relatively small groups of participants may be able to predict population level responses to media. Future research is needed to examine the boundary conditions of the effects observed, and to elucidate the complex inputs that comprise this value signal that seems to be integrated by VMPFC.

FUTURE DIRECTIONS

Over the past decade, our understanding of the neural systems that support evaluation, preferences, attitudes, and persuasion has grown into a base that will support ongoing investigations. Future investigations will continue to explore the ways in which the brain generates evaluations of the social

environment, makes judgments, forms preferences, and acts upon these attitudes and preferences under various circumstances.

Behavioral research has clearly demonstrated that the dynamic process of evaluation and attitude change differs depending on factors such as the initial strength of attitudes, and factors related to the cause of potential attitude change (Eagly & Chaiken, 1993; Petty & Cacioppo, 1986). Future work is needed to explore moderators and boundary conditions of the neural bases of each attitudinal process described in this chapter. Research is also needed to explore the relationship between the neurocognitive predictors of attitudes, intentions, and behaviors under different circumstances. For example, whereas initial steps have been taken to explore neural predictors of behavior change following a persuasive message, and to understand how this pattern differs depending on the behavior in question, it will be of interest to more specifically interrogate factors related to the message, communicator, and message delivery.

Lastly, as a final caveat, neuroimaging allows examination of multiple processes in concert, and may allow us to link our understanding of human psychological processes to a vast neuroscience literature in animal models; however, it is also subject to inherent weaknesses. For example, the scanner environment is likely to reduce our ability to simulate real-life situations, and may also introduce confounds related to the novelty of the situation and/or the conditions under which information is delivered. Likewise, we must be cautious in our use of reverse inference (Poldrack, 2006); neuroimaging research can inspire novel hypotheses, however these hypotheses must be tested. Thus, just as behavioral research informs the questions that are asked in fMRI, it will be useful to consider novel hypotheses generated by the work that can be tested outside of the scanner in a more naturalistic environment, and to link neuroimaging findings to real-world and longitudinal outcomes. By employing an iterative process in which behavioral and neuroimaging research continues to inform one another, both disciplines will benefit.

NOTE

1 Self-presentation may be of concern in terms of how one appears to others, but may also arise to the extent that individuals are motivated to view themselves as unprejudiced.

REFERENCES

Aaker, J. L., & Williams, P. (1998). Empathy versus pride: The influence of emotional appeals across cultures. *Journal of Consumer Research, 25*(3), 241–261.

Albarracin, D., Johnson, B. T., & Zanna, M. P. (eds.) (2005). *The Handbook of Attitudes*. Mahwah, NJ: Lawrence Erlbaum.

Allport, G. W. (1935). Attitudes. In C. M. Murchison (ed.) *Handbook of Social Psychology*. Winchester, MA: Clark University Press.

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. *Psychol Sci, 15*(2), 88–93.

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

Amodio, D., & Lieberman, M. (2009). Pictures in our heads: Contributions of fMRI to the study of prejudice and stereotyping. In T. Nelson (ed.) *Handbook of Prejudice and Discrimination*. New York: Taylor & Francis.

Aristotle (1924/1954). The rhetoric of Aristotle. *Aristotle, with an English Translation: The "Art" of Rhetoric, by John Henry Freese*. Oxford: Clarendon Press.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition, 21*(1), 61–87.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *J Pers Soc Psychol, 62*(6), 893–912.

Berns, G. S., & Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology, 22*, 154–160.

Bishop, G. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly, 44*(2), 198.

Bishop, G. D., Hamilton, D. L., & McConahay, J. B. (1980). Attitudes and nonattitudes in the belief systems of mass publics. *The Journal of Social Psychology, 110*, 53–64.

Bruneau, E. G., & Saxe, R. (2010). Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *Neuroimage, 52*(4), 1704–1711.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci, 12*(1), 1–47.

Campbell, R. G., & Babrow, A. S. (2004). The role of empathy in responses to persuasive risk communication: Overcoming resistance to HIV prevention messages. *Health Commun, 16*(2), 159–182.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*(5), 541–560.

Chua, H., Liberzon, I., Welsh, R., & Strecher, V. (2009a). Neural correlates of message tailoring and self-relatedness in smoking cessation programming. *Biol Psychiatry, 65*(2), 165–168.

Chua, H., Polk, T., Welsh, R., & Liberzon, I. (2009b). Neural responses to elements of a web-based smoking cessation program. *Stud Health Technol Inform, 144*, 174–178.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufte (ed.) *The Quantitative Analysis of Social Problems*. Reading, MA: Addison-Wesley.

Cunningham, W., Johnson, M., Gatenby, J., Gore, J., & Banaji, M. (2003). Neural components of social evaluation. *J Pers Soc Psychol, 85*, 639–649.

Cunningham, W., Johnson, M., Raye, C., Chris Gatenby, J., Gore, J., & Banaji, M. (2004a). Separable neural components in the processing of black and white aces. *Psychol Sci, 15*(12), 806–813.

Cunningham, W., Raye, C. L., & Johnson, M. K. (2004b). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J Cogn Neurosci, 16*(10), 1717–1729.

Cunningham, W., & Zelazo, P. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11*(3), 97–104.

Cunningham, W., Zelazo, P., Packer, D. J., & Van Bavel, J. J. (2007). The Iterative Reprocessing Model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25*(5), 736–760.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Reviews in Neuroscience, 15*(1), 353–375.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *J Pers Soc Psychol, 56*(1), 5–18.

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Soc Cogn Affect Neurosci, 2*(3), 189–198.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *J Pers Soc Psychol, 82*(1), 62–68.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *J Exp Psychol Gen, 127*(3), 286–303.

Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Eagly, A. H., & Chaiken, S. (2005). *Attitude Research in the 21st Century: The Current State of Knowledge*. Mahwah, NJ: Lawrence Erlbaum.

Eagly, A., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition, 25*(5), 582–602.

Eberhardt, J. (2005). Imaging race. *American Psychologist, 60*(2), 181–190.

Falk, E. B. (2010). Communication neuroscience as a tool for health psychologists. *Health Psychology, 29*(4), 346–354.

Falk, E. B., Berkman, E. T., & Lieberman, M. D. (2012). From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychol Sci, 23*(5), 439–445, online, doi: 10.1177/0956797611434964.

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010a). Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience, 30*, 8421–8424.

Falk, E. B., Berkman, E. T., Whalen, D., & Lieberman, M. D. (2011). Neural activity during health messaging predicts reductions in smoking above and beyond self-report. *Health Psychology, 30*, 177–185.

Falk, E., Rameson, L., Berkman, E., Liao, B., Kang, Y., Inagaki, T. K., & Lieberman, M. (2010b). The neural correlates of persuasion: A common network across cultures and media. *J Cogn Neurosci, 22*, 2447–2459.

Falk, E. B., Welborn, L., Morelli, S., Dambacher, K., & Lieberman, M. D. (under review). The neuroscience of buzz: Neural correlates of message propagation.

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? *J Pers Soc Psychol, 69*(6), 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *J Pers Soc Psychol, 50*, 229–238.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson & Company.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci, 358*(1431), 459–473.

Gawronski, B. (2007). Editorial: Attitudes can be measured! But what is an attitude? *Social Cognition, 25*(5), 573–581.

Gozzi, M., Raymont, V., Solomon, J., Koenigs, M., & Grafman, J. (2009). Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia, 47*(10), 2125–2132.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646–648.

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Soc Cogn Affect Neurosci, 2*(1), 45–51.

Hart, A., Whalen, P., Shin, L., McInerney, S., Fischer, H., & Rauch, S. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroReport, 11*(11), 2351.

Hovland, C. I. (1949). Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist, 14*, 8–17.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and Persuasion: Psychological Studies of Opinion Change*. New Haven, CT: Yale University Press.

Jacobsen, T., Schubotz, R., Hofel, L., & Cramon, D. (2005). Brain correlates of aesthetic judgment of beauty. *Neuroimage, 10*, 276–285.

Jarcho, J. M., Berkman, E. T., & Lieberman, M. D. (2011). The neural basis of rationalization: Cognitive dissonance reduction during decision-making. *Soc Cogn Affect Neurosci, 6*, 460–467.

Kaplan, J., Freedman, J., & Iacoboni, M. (2007). Us versus them: Political attitudes and party affiliation influence neural response to faces of presidential candidates. *Neuropsychologia, 45*(1), 55–64.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *J Pers Soc Psychol, 81*(5), 774–788.

Klucharev, V., Smidts, A., & Fernandez, G. (2008). Brain mechanisms of persuasion: How "expert power" modulates memory and attitudes. *Soc Cogn Affect Neurosci, 3*(4), 353–366.

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron, 53*(1), 147–156.

Knutson, K., Wood, J., Spampinato, M., & Grafman, J. (2006). Politics on the brain: An fMRI investigation. *PSNS, 1*(1), 25–40.

Koenigs, M., & Tranel, D. (2008). Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Soc Cogn Affect Neurosci, 3*(1), 1–6.

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

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracin, B. T. Johnson, & M. P. Zanna (eds.) *The Handbook of Attitudes*. Mahwah, NJ: Lawrence Erlbaum.

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

Lieberman, M. (2010). Social cognitive neuroscience. In S. Fiske, D. Gilbert, & G. Lindzey (eds.) *Handbook of Social Psychology* (5th edn). New York: McGraw-Hill.

Lieberman, M., Hariri, A., Jarcho, J., Eisenberger, N., & Bookheimer, S. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nat Neurosci, 8*(6), 720–722.

Lieberman, M., Ochsner, K., Gilbert, D., & Schacter, D. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychol Sci, 12*(2), 135–140.

Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology, 38*, 405–413.

Mason, M. F., Dyer, R. G., & Norton, M. I. (2009). Neural mechanisms of social influence. *Organizational Behavior and Human Decision Processes, 110*(2), 152–159.

McClure, S. M., Li, J., Tomlin, D., Cypert, K., Montague, L., & Montague, P. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*(2), 379–387.

McConahay, J. P. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (eds.) *Prejudice, Discrimination, and Racism*. Orlando, FL: Academic Press.

Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J Neurosci, 21*(12), RC150.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101–115.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist, 56*(9), 717–734.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.

Petty, R. E., Priester, J. R., & Wegener, D. T. (1994). Cognitive processes in attitude change. In R. S. J. Wyer (ed.) *Handbook of Social Cognition* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum.

Phelps, E. (2001). Faces and races in the brain. *Nat Neurosci, 4*(8), 775–776.

Phelps, E., Cannistraci, C. J., & Cunningham, W. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia, 41*(2), 203–208.

Phelps, E., O'Connor, K. J., Cunningham, W., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activity. *J Cogn Neurosci, 12*(5), 1–10.

Phelps, E., & Thomas, L. (2003). Race, behavior, and the brain: The role of neuroimaging in understanding complex social behaviors. *Political Psychology, 24*(4), 747–758.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *J Pers Soc Psychol, 75*, 811–832.

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci, 27*(37), 9984–9988.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proc Natl Acad Sci U S A, 105*(3), 1050–1054.

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci, 10*(2), 59–63.

Quadflieg, S., Turk, D. J., Waiter, G. D., Mitchell, J. P., Jenkins, A. C., & Macrae, C. N. (2009). Exploring the neural correlates of social stereotyping. *J Cogn Neurosci, 21*(8), 1560–1570.

Richeson, J., Baird, A., Gordon, H., Heatherton, T., Wyland, C., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nat Neurosci, 6*(12), 1323–1328.

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z.-L., Nandy, A., & Maddox, K. B. (2007). The effects of skin tone on race-related amygdala activity: An fMRI investigation. *Soc Cogn Affect Neurosci, 2*(1), 39–44.

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Z Exp Psychol, 48*(2), 94–106.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *J Exp Psychol Gen, 133*(2), 139–165.

Sharot, T., De Martino, B., & Dolan, R. J. (2009). How choice reveals and shapes expected hedonic outcome. *J Neurosci, 29*(12), 3760–3765.

Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychol Sci, 16*(5), 397–402.

Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (eds.) *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception*. San Diego, CA: Academic Press.

Tusche, A., Bode, S., & Haynes, J. D. (2010). Neural responses to unattended products predict later consumer choices. *J Neurosci, 30*(23), 8024–8031.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychol Sci, 19*(11), 1131–1139.

van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nat Neurosci, 12*(11), 1469–1474.

Vartanian, O., & Goel, V. (2004). Neuroanatomical correlates of aesthetic preference for paintings. *NeuroReport, 15*(5), 893.

Walker, P. M., Silvert, L., Hewstone, M., & Nobre, A. C. (2008). Social contact and other-race face processing in the human brain. *Soc Cogn Affect Neurosci, 3*(1), 16–25.

Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An FMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election. *J Cogn Neurosci, 18*(11), 1947–1958.

Wheeler, M., & Fiske, S. (2005). Social-cognitive goals affect amygdala and stereotype activation. *Psychol Sci, 16*(1), 56–63.

Zajonc, R. B., & Markus, H. (1982). Affective and cognitive factors in preferences. *Journal of Consumer Research, 9*(2), 123–131.

Zamboni, G., Gozzi, M., Krueger, F., Duhamel, J. R., Sirigu, A., & Grafman, J. (2009).

Individualism, conservatism, and radicalism as criteria for processing political beliefs: A parametric fMRI study. *Soc Neurosci, 4*(5), 367–383.
Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *Neuroimage, 15*(4), 983–991.

CHAPTER FIVE

Interpersonal trust as a dynamic belief

Ewart de Visser and Frank Krueger

INTRODUCTION

Trust pervades nearly every social aspect of our daily lives, from personal relationships to organizational interactions encompassing social, economic, and political exchange. There is a vast literature on interpersonal trust that has examined the phenomenon from several academic disciplines including psychology, economics, political science, evolutionary biology, and neuroscience; however, the underlying neural architecture of interpersonal trust is still not well understood. In this chapter, we sketch out an integrative cognitive neuroscience framework to understand how interpersonal trust emerges as a dynamic belief from the interplay of specific complementary brain circuits. By drawing the recent findings in the field of cognitive neuroscience together into a coherent picture, one might gain a better understanding of the underlying dynamic neural architecture of trust, which operates within the immediate spheres of nature and nurture and determines which forms of social, economic, and political institutions develop within social groups.

First, we provide a working definition of interpersonal trust and describe how it can be empirically measured. Second, we introduce the Motivation-Affect-Cognition (MAC) model of interpersonal trust, in which trust emerges through the interplay of three specific complementary systems: (1) a cognitive system, (2) a motivational system, and (3) an affective system. Third, we review the different lines of evidence supporting the underlying neural architecture of the MAC model, focusing on both functional neuroimaging and neuropsychological brain lesion studies. Fourth, we describe how the dynamic neural architecture of interpersonal trust is modulated by oxytocin, a peptide that functions both as a hormone and a neurotransmitter broadly influencing affiliative behavior. Finally, we will close the chapter by pointing to open

CONTEMPORARY TOPICS IN COGNITIVE NEUROSCIENCE SERIES

Reflecting contemporary and controversial issues in the study of cognitive neuroscience, the series aims to present a multi-disciplinary forum for cutting edge debate that will help shape this burgeoning discipline. It offers leading figures in the field and the best new researchers an opportunity to showcase their own work, expand on their own theories and place these in the wider context of the field.

Titles in the series may be authored or edited; each book must aim to make a contribution to a specific topic by reviewing and synthesizing the existing research literature, by advancing theory in the area, or by some combination of these missions.

Published titles in the series

Neuroscience of Decision Making
Edited by Oshin Vartanian & David. R. Mandel

The Neural Basis of Human Belief Systems
Edited by Frank Krueger & Jordan Grafman

Language and Action in Cognitive Neuroscience
Edited by Yann Coello & Angela Bartolo

For more information about the series, please visit **www.psypress.com/ctcn**

The Neural Basis of Human Belief Systems

Edited by Frank Krueger & Jordan Grafman

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK