Neuroscience Approaches in Social and Personality Psychology

David M. Amodio¹, Eddie Harmon-Jones², and Elliot T. Berkman³

¹University of Amsterdam, New York University, ²The University of New South Wales, and ³University of Oregon

Abstract
Social neuroscience is an interdisciplinary approach to studying the social mind and behavior by considering their neural and biological underpinnings. In this chapter, we describe the methodological approach of social neuroscience and review research that has applied this approach to address the interplay of the person and situation in the domains of social cognition, attitudes, emotion and motivation, intergroup relations, and personality. We provide critical discussion of how neuroscience may contribute to classic questions in personality and social psychology, and we describe how the social neuroscience approach promotes the integration of dispositional and situational accounts of the mind and behavior.

Keywords: social, personality, cognitive, neuroscience, emotion, attitudes, intergroup, prejudice, review

Contact information:
David M. Amodio
Department of Psychology and Center for Neural Science
New York University
6 Washington Place, rm. 760
New York, NY 10003
david.amodio@nyu.edu
Social neuroscience 2

Neuroscience Approaches to Social and Personality Psychology

The contemporary interest in neural and biological correlates of social and personality processes may seem new, but it is actually a return to form for the field of social and personality psychology. Founding psychologists such as James and Wundt were trained in physiology, and they approached issues of thought, emotion, memory, and perception with a firm belief that mental processes were rooted in the brain. To these psychologists, the mind and the brain were symbiotic, and a consideration of neural processes, along with dispositional and interpersonal processes, provided a natural and comprehensive approach to understanding the mind. Early psychological research on social processes, such as bluffing to other players during a poker game (Riddle, 1925) or responding to socially taboo words and phrases (Darrow, 1929), incorporated psychophysiological measures to complement behavioral observations. Indeed, early research on intergroup interactions and prejudice used physiological measures to examine anxiety during an interracial encounter (Rankin & Campbell, 1955). Hence, physiological measures have long served as an important component of an integrated approach to social-personality psychology.

As the field of psychology developed, a shift toward behaviorism effectively vanquished the role of the mind and brain from the study of behavior, and the field of neuroscience branched away from psychology. While research on neural function continued at lower levels of analysis (e.g., cellular and molecular), it no longer made contact with the higher-level processes of social and personality psychology. When a focus on mental function reemerged during the cognitive revolution, it was generally separate from neural and biological function (with some notable exceptions, e.g., Cacioppo & Petty, 1983; Frith, Morton, & Leslie, 1991; Gazzaniga, 1985). However, with recent advances in brain imaging technology, research on cognitive neuroscience has surged over the past two decades, making increasing contact with questions of social cognition (Amodio & Ratner, 2013; Ochsner & Lieberman, 2001). In this way, social and personality
psychologists have begun to reintegrate neural and biological approaches into the range of methods used to understand the social mind and behavior.

In this chapter, we review the contemporary social neuroscience approach to social-personality psychology. This general approach began to appear with increasing frequency in the laboratories of social psychologists, cognitive neuroscientists, and neurologists during the 1990s. The term “social neuroscience” was coined in print by Cacioppo and Berntson (1992) and tends to describe the broad enterprise of examining the interplay of social and physiological levels of analysis. More recent formulations by Ochsner and Lieberman (2001) and Klein and Kihlstrom (1998) incorporated contemporary ideas from cognitive neuroscience and neuropsychological patient literatures, respectively, prompting new aspects of social neuroscience referred to as “social cognitive neuroscience” and “social neuropsychology.” Over the past decade, social neuroscience has been the subject of several dedicated research conferences, culminating in the formation of the Social and Affective Neuroscience Society in 2008 and the Society for Social Neuroscience in 2010. Whereas social neuroscience was seen as a novelty at social-personality meetings merely a decade ago, it is now fully integrated into the social/personality psychologist’s methodological repertoire.

The present volume highlights the interplay of social and personality factors in studies of the mind and behavior. The social neuroscience approach fits this theme well. Integrative at its core, social neuroscience encompasses the study of personality and individual differences as well as situational and environmental effects, as they interface with cognitive processes and neural mechanisms. In this chapter, we use the term social neuroscience to refer to this general integrative approach, although it might just as easily be called “social-personality neuroscience.”

In what follows, we begin with an overview of the methodological approaches used in social neuroscience. We then describe social neuroscience research across major areas of social and
personality psychology, with a focus on how neuroscience and physiological approaches pertain to social-personality theory.

**The social neuroscience approach**

Broadly speaking, social neuroscience refers to an integrative approach that can be applied to any scientific question concerning social processes and the brain. However, the types of questions that are addressed with this approach vary considerably across disciplines. To the social psychologist, *social neuroscience* refers to an interdisciplinary research approach that integrates theories and methods of neuroscience (and other biological fields) to address social psychological questions. To a cognitive neuroscientist, it often refers to research that addresses questions about the neural substrates of social processes, with a focus on understanding neural function. To an animal behaviorist, social neuroscience research may address questions about the neural, hormonal, and genetic mechanisms associated with low-level social behaviors, such as dominance and affiliation. And to a behavioral economist, the social neuroscience approach might be applied to understand the neurocognitive underpinnings of economic choices and negotiations, as well as the biases that often emerge in human decision making. Thus, many “social neuroscience” studies examine questions outside the typical purview of social-personality psychology, and therefore it is helpful for consumers of this literature to carefully consider the question asked by a particular study. In this section, we describe the two main types of questions asked in human social neuroscience and their corresponding methodological approaches. We also describe the critical role of reverse inference in drawing conclusions from neuroscience findings.

**Brain mapping approach**

Brain mapping studies ask “*where in the brain is _____?*” For example, where in the brain is fear? Where is episodic memory? Where is love? Where is the self? Human brain mapping is a
cornerstone of modern cognitive neuroscience. It concerns the mapping of basic psychological processes to particular regions of the brain.

Early forms of brain mapping involved the probing of exposed brain tissue by a neurosurgeon while a patient reported his or her experience. Today, relatively non-invasive neuroimaging measures, such as functional magnetic resonance imaging (fMRI), are often used for a similar purpose. In cognitive neuroscience, this approach is used to map relatively low-level psychological processes such as basic forms of sensation, perception, and specific aspects of learning and memory. As a general rule, lower-level cognitive processes can be mapped more directly onto specific physiological responses than more complex high-level processes.

In social psychology, researchers have attempted to map very high-level psychological processes, such as social emotions, the self-concept, trait impressions, and political attitudes, onto the brain as well. This is where things get trickier. For example, to study the neural basis of romantic love, a researcher might scan participants’ brains while they view pictures of their significant others in comparison to strangers. Similarly, to study the neural basis of the self, one might scan the brain while subjects judge whether trait adjectives described themselves or another person. Thus, such studies would apply the same logic to identifying the neural substrates of very high-level processes as neuroscientists have applied in the neural mapping of very low-level processes, such as edge detection in vision. Although high-level psychological ascriptions of brain activity may have heuristic value, they may risk obscuring the important lower-level mechanisms that the observed brain activations likely represent.

A defining feature of the brain mapping approach is that it seeks to create a valid mapping of psychological processes onto a pattern of neurophysiological responses. Pure brain-mapping studies are undertaken with few prior assumptions about the psychological function of a brain region – indeed, the point of such studies is to establish ideas about function through the process of induction
across multiple studies using a variety of conceptually-similar tasks and manipulations. This approach is potentially useful for generating new ideas about commonalities in the cognitive processes that may underlie two otherwise distinct psychological functions. For example, researchers have observed that down-regulation of both negative emotions and cravings for appealing but unhealthy foods activate a common set of regions, and thus concluded that regulation of aversive and appetitive responses share neurocognitive mechanisms and might benefit from the same kinds of interventions (Giuliani & Berkman, 2015). Although this approach does not tell us exactly how or why they might be related, since the true function of the neural activity is difficult to discern, it nevertheless provokes new ideas about potential relationships between psychological processes. Importantly, brain mapping studies are useful for probing the psychological function of neural regions, but they are usually not appropriate for testing hypotheses about the relationship between two psychological variables or the effects of an experimental manipulation on a psychological variable.

*Hypothesis testing approach*

The hypothesis testing approach in social neuroscience is used to test relationships between psychological variables. This approach begins with the assumption that a particular brain region reflects a specific psychological process. In this regard, it does not concern brain mapping, but instead relies on past brain-mapping studies to have already established the validity of neural indicators. For example, a social psychologist who studies intergroup prejudice might hypothesize that implicit racial bias is rooted in mechanisms of classical fear conditioning (Amodio, Harmon-Jones, & Devine, 2003). To test this hypothesis, one might measure brain activity in the amygdala (Fig 1) – a structure implicated in fear conditioning in many studies – while a participant completes a behavioral measure of implicit racial bias. In this case, the construct validity of the neural measure of fear conditioning (amygdala activity) is already reasonably established (e.g., through extensive...
multi-method experimentation, or though large-scale meta-analysis; Yarkoni et al., 2011), and the question concerns not the meaning of brain activations, but experimental effects among psychological variables. It is the hypothesis testing approach of social neuroscience that is of primary interest to social-personality psychologists. Whereas brain-mapping studies typically inform our understanding of the brain, hypothesis-testing studies can inform psychological theories of the mind.

Critically, brain mapping and psychological hypothesis testing approaches should not be combined within a single analysis; major inferential problems occur as a result (Amodio, 2010a). This is because a test of a psychological hypothesis assumes that the mapping of a psychological variable to a neural structure is already established (e.g., that the neural measure has construct validity). But when these approaches are combined, there is a risk of defining the neural operationalization of a psychological construct on the basis of whether it supports one’s theoretical hypothesis.

To illustrate, imagine that a researcher wants to test the hypothesis that empathy involves self-reflection. He does so by examining the correlation of self-reported empathy with activity in the medial prefrontal cortex (mPFC, Fig 2), a region linked to self-related processing in some previous studies, while participants view images of people in pain. However, the problem with this approach is that the mPFC has been associated with several other processes besides self-reflection, and thus its construct validity as an indicator of “the self” is difficult to establish. If the research found an area of mPFC to correlate with empathy ratings, he would simultaneously be inferring that this area must be the “self” region (establishing the construct) and that “self” activity is indeed associated with greater empathy (testing the psychological hypothesis). This confounds the discrete steps of establishing construct validity (brain mapping) and internal validity (hypothesis testing). Because the construct is validated on the basis of the hypothesis-testing correlation analysis, the logic of the
test is circular (Amodio, 2010a). That is, the meaning of a region is validated only after its activation supports the researcher’s hypothesis. This analytical approach is clearly problematic, yet it arises frequently in social neuroscience research, primarily because the social psychological processes of interest are complex and difficult to localize precisely in the brain. Consumers of social/personality neuroscience should be aware of such practices and cautious of their use.

Reverse inference

When considering the two general approaches described above, the issue of reverse inference is often a concern. Reverse inference refers to a form of reasoning used heavily in social and cognitive neuroscience to infer the psychological meaning of a brain activation based on previous findings (Poldrack, 2006). In studies of brain mapping, a psychological process is manipulated and the resulting pattern of brain activity is observed. The inference that the psychological manipulation produced the brain activity may be described as a forward inference, in that the brain activity clearly follows from the manipulation. The inference is based on the known validity of the manipulation. By contrast, the inference of a psychological process from an observed pattern of brain activity is a reverse inference. In this case, the precise meaning of the brain activation is ambiguous and inferred from other studies that have used a particular manipulation to activate the same area. The practice of reverse inference becomes increasingly problematic to the extent that the source of inference – in this case, a brain activation – could reflect different psychological processes (Cacioppo et al., 2003; Poldrack, 2006). In studies of low-level vision, reverse inference is a comparatively lesser problem (but still a serious issue). For example, retinotopic mapping of stimuli onto primary visual cortex provides a relatively valid index of basic visual perception. But as psychological variables become more complex, as they do with cognitive and social processes, the mapping between a particular brain region and a psychological process becomes less certain. In these cases, reverse inference can be a serious problem.
The truth is that virtually all cognitive and social neuroscience studies rely on reverse inference. That is, to the extent that a neural activation is interpreted as reflecting a psychological process, the use of reverse inference is unavoidable. However, researchers can take steps to bolster the strength of a reverse psychological inference by enhancing the construct validity of a neural indicator and the strength of their experimental designs, such as through the careful use of theory, converging evidence from other studies (including animal research), and the use of behavioral tasks that provide valid manipulations of a construct and interpretable behavioral data.

Serious efforts have been made in recent years to estimate the validity of reverse inferences. These approaches begin by acknowledging that the essential problem with drawing inferences about psychological process from neural activity is selectivity: each brain region is associated with many processes, and the baseline likelihood that each of process will occur is unknowable from any one data set. A solution to this problem is to estimate the probability that a region is active (known as the prior in Bayesian terms) and the probability that it is active given a particular mental process (known as the posterior) by aggregating information across many studies. Doing so allows researchers to quantify the strength of a reverse inference. A first pass at this kind of large-scale Bayesian reverse-inference has been implemented in a web-based tool called Neurosynth (Yarkoni et al., 2011). Social neuroscience researchers have begun to use Neurosynth to leverage prior knowledge to more precisely quantify the likelihood of reverse inferences in a single study. Meta-analytic tools, more generally, are a highly promising avenue across the neurosciences because they can address many problems inherent in neuroimaging research such as inadequate power due to limitations on sample sizes.

What types of social and personality questions are amenable to a neuroscience analysis?

First and foremost, the brain is a mechanism, and an extremely complex one at that. Hence, neuroscience models and methods are useful for the study of psychological mechanisms, such as
those involved in action control, perception, and attention. Psychological phenomena that are not mechanisms, but correspond more closely to appraisals (e.g., attitudes and beliefs), subjective psychological states, abstract psychological structures (e.g., the self), and high-level representation, may be less amenable to a neuroscience level of analysis. For a social/personality psychologist who is considering the potential benefit of a neuroscience approach, the most critical issue is whether one’s question concerns basic psychological mechanism. Can the components of one’s mechanistic model be described in terms of low-level functions, such as perception, sensation, and lower-level aspects of cognition and motivation? If so, then neuroscience models may be particularly useful. If the psychological phenomena of interest cannot be conceptualized at these lower levels of analysis, but rather are most meaningful at a high level of construal (e.g., the self), then it may be more difficult to make valid inferential connections between psychological theory and the brain.

**Methods of social and personality neuroscience**

Contemporary social neuroscience makes use of a wide range of methods that are often used in combination with the more traditional tools of personality and social psychology. In addition to new technologies for measurement, social neuroscience methodology relies on the use of experimental designs, valid manipulations of psychological states and processes, and careful inference and interpretation. Here, we describe the most prominent methods currently used in the field and briefly discuss their relative advantages as they relate to experimental designs, issues of construct validity, and psychological inference. A more detailed description of methods in neuroscience approaches to social and personality psychology is provided by Harmon-Jones and Beer (2009), and recent discussions of inference and validity in social neuroscience can be found in Amodio (2010a), and Cacioppo et al. (2003).

Early studies taking a social neuroscience approach primarily used peripheral physiological methods, such as electrocardiogram (e.g., heart rate), galvanic skin response (i.e., skin conductance,
a measure of sympathetic activation (vis-à-vis palm sweating), and electromyography (e.g., measures of facial muscle activity related to emotional expressions).

More recently, with a greater focus on neural function, researchers have increasingly utilized neuroimaging methods. These include functional magnetic resonance imaging (fMRI), which measures the flow of oxygenated blood in the brain, and electroencephalography (EEG), which measures electrical activity produced from the firing of neuron populations, as well as some less-frequently used methods. Event-related potentials (ERPs) represent a burst of EEG activity in response to a discrete event, such as a stimulus or subject response (Amodio & Bartholow, 2011). fMRI yields high spatial resolution and thus is optimal for determining the location of activity within the brain (Berkman, Cunningham, & Lieberman, 2014). But because it assesses slow-moving blood flow, its temporal resolution is relatively poor. By contrast, EEG/ERP yields high temporal resolution and is thus optimal for assessing the timing of a neural process, but its spatial resolution is comparatively poor. Given their relative strengths, researchers may select fMRI or EEG methods to suit their question, or use both approaches in complementary studies within a program of research.

Neuroimaging and psychophysiological approaches may also be combined with measures of hormones, immune factors, and DNA, for example, to provide convergent evidence for a physiological process of interest. There are also increasingly sophisticated ways to increase or decrease brain activity in a focal region with neurostimulation techniques such as transcranial direct-current stimulation (tDCS) and transcranial magnetic stimulation (TMS). The modest spatial resolution of these techniques is offset by a huge advantage: by manipulating brain activity, they allow for more direct inferences about the role of a region in a particular psychological process. However, as with traditional methods in social-personality psychology, the utility of these measures depends on the quality of the question, the experimental paradigm, and careful interpretation.
Major content areas of social/personality neuroscience

In this section, we provide a broad review of the social neuroscience literature in the areas of social cognition and the self, attitudes, emotion and motivation, intergroup relations, and individual differences. Although our review distinguishes these five areas of research for convenience, their content overlaps substantially. In each area, we will integrate methods and levels of analysis, and we will discuss how research in each area has contributed to social/personality psychology theories.

Social Cognition and the Self

Neuroscience research has sought to expand our understanding of the self and social cognition by identifying specific neural structures and networks that underlie complex social behavior (Amodio & Ratner, 2011a; Lieberman, 2007; Frith & Frith, 2012). Moreover, through cross-species comparisons in neural structure and capacities, it also addresses questions of whether aspects of the self and social cognition are unique to humans (Farah & Gillian, 2005; Rushworth, Mars, & Sallet, 2013). In this section, we provide an overview on this area of research as it related to major topics within social-personality psychology.

Automatic and controlled processing

Theories of automatic and controlled processes are foundational to modern social cognition. Mechanisms of automaticity and control also constitute a central topic in cognitive psychology, and a large body of cognitive neuroscience research has been devoted to their elucidation. These processes have been studied primarily in two different broad research literatures on memory and cognitive control.

Automaticity. Research in the memory literature is highly relevant to research on automaticity, as it suggests different forms of automaticity with regard to its role in cognition and behavior. In this literature, researchers often distinguish between neural correlates of explicit (declarative) and implicit (non-declarative) memory (Squire & Zola, 1996). Although implicit and
automatic refer to different properties of a process (i.e., degree of awareness vs. degree of intentionality), implicit forms of memory, such as classical fear conditioning and procedural memory (i.e., skill or habit learning), have important automatic characteristics and thus are relevant to the present discussion. Studies of non-human animals have identified the amygdala as a critical structure in the learning and expression of fear conditioning (LeDoux et al., 1988), a role that has been corroborated in fMRI studies of fear conditioning in humans (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998). Fear conditioning and expression in the amygdala may occur rapidly and without conscious mediation, suggesting it underlies one form of automaticity. However, because fear conditioning reflects a passive, reactive response to a learning threat, it does not map on cleanly to the traditional concept of an “automatic” process, in the sense that automaticity refers to an active behavioral or cognitive process.

Research on procedural memory suggests a different form of automaticity with more direct connections to action. This work has focused on the basal ganglia, a set of interconnected structures that include the striatum (Fig 3; caudate, putamen, and nucleus accumbens), globus pallidus, and their dopaminergic inputs from the midbrain (substantia nigra and ventral tegmentum). Research has identified regions of basal ganglia as being critical for implicit skill learning, such as when subjects learn response associations that are embedded implicitly in a task procedure (Foerde, Knowlton, & Poldrack, 2006; Squire & Zola, 1996). This form of learning and memory more closely resembles the type of automatic processes studied in the social and personality psychology literature. It is notable that, in recent years, studies of economic decisions and reward learning have also focused on the role of basal ganglia, but as a substrate for reward prediction error (e.g., Diuk, Tsai, Wallis, Botvinick, & Niv, 2013). However, given the clear interrelatedness of skill learning, motor output, and reward, it is likely that the basal ganglia are centrally involved in implicit learning related to goal driven behaviors (including automatic responses).
In the social cognition literature, automaticity is often demonstrated in sequential priming tasks, when a prime word is shown to facilitate the categorization of an associated target word. This idea was originally adapted from cognitive psychology research examining semantic associations between words, such as “bread-butter” or “doctor-nurse” (Meyer & Schaneveldt, 1971). Research on the neural processes associated with semantic priming has generally found evidence of activations in left posterior PFC (e.g., Demb et al., 1995; Wagner, Gabrieli, & Verfaellie, 1997) and temporal cortex (Schacter & Buckner, 1998), and deactivations in regions linked to attention (e.g., in the parietal cortex, Fig 3; Gabrieli, 1998). Given other research implicating the left PFC in approach-related motivation and action tendencies (Harmon-Jones, Gable, & Peterson, 2010), this pattern of neural correlates suggests a link between automatic semantic processes and goal-driven behavior (Amodio, 2008), consistent with the idea that “thinking is for doing”.

Together, the neuroscience research on fear-conditioning, procedural memory and reward, and semantic associations has helped to distinguish different types of automatic processes. In doing so, these studies clarify the functions of automaticity and shed new light on how automatic processes operate and change. These advances have already inspired new theories of implicit social cognition (e.g., Amodio & Ratner, 2011a; Hackel, Doll, Amodio, 2015).

**Control.** Cognitive neuroscience research on controlled processing has begun to elucidate key components of the regulatory processes as they relate to social-personality processes. For example, this work has examined subjects’ brain activity while they completed classic cognitive control tasks, such as the Stroop color naming task, the Eriksen Flankers Task, or the Go/No-Go task. Each of these tasks includes two main types of trials that require either a high or low degree of control. In one set of trials, subjects make responses that are facilitated by either existing semantic associations (e.g., the color-naming Stroop task), perceptual cues (e.g., the Eriksen Flankers task), or expectancies (e.g., the Go/No-Go). Responses on these trials are thought to benefit from
automatic processing. On other trials, the participant must override the automatic influence or otherwise bias behavior toward the alternative in order to deliver the correct task response. When brain activity during the “control” trials is compared with activity to “automatic” trials, several regions of the anterior cingulate cortex (ACC) and PFC are typically activated.

In particular, the ACC responds to instances when a prepotent, or automatic, response is inconsistent with the task goal, such as when the text of a color word interferes with one’s goal to name the ink color (Carter et al., 1998). Botvinick, Braver, Barch, Carter, and Cohen (2001) proposed that the ACC provides a conflict monitoring function, such that it is involved in detecting conflict between alternative response tendencies and, when conflict arises, it signals regions of the PFC involved in implementing one’s intended response over other tendencies.

Consistent with the conflict monitoring hypothesis, anatomical research on monkeys has revealed that the ACC is strongly interconnected with motor structures as well as PFC regions associated with high-level representations of goals and actions (Miller & Cohen, 2001). Thus, this research reveals that control involves at least two different components—conflict monitoring and control implementation—which in turn may help to explain aspects of behavior, social cognition, and individual differences.

Finally, several studies have shown that conflict-related ACC activity occurs in the absence of awareness (Berns, Cohen, & Mintun, 1997; Nieuwenhuis et al., 2001), suggesting that it represents a preconscious component of control that initiates the engagement of more deliberative components. The findings of cognitive neuroscience research on conflict monitoring and control have been applied to test theories of self-regulation in social-personality contexts such as stereotyping and prejudice (Amodio et al., 2004; Amodio, Harmon-Jones, & Devine, 2008; Bartholow, Dickter, & Sestir, 2006), individual differences related to anxiety and neuroticism (Amodio, Master, Yee, & Taylor, 2008), social anxiety (Schmid, Kleiman, & Amodio, 2015a),
political orientation (Amodio, Jost, Master, & Yee, 2007), social power (Schmid, Kleiman, & Amodio, 2015b), religiosity (Inzlicht et al., 2009), and social exclusion (Eisenberger et al., 2003).

Although several regions of the PFC are activated during attempts at response control, a general pattern has been observed in the literature whereby left-sided PFC regions are associated with the implementation of intended actions, whereas right-sided PFC activity has been associated with the intentional inhibition of action (Aron, 2007). It is notable that neuroimaging research is almost always conducted on right-handed participants, and observed hemispheric asymmetries in patterns of control are likely related to the lateral specialization of hand (and foot) dominance (Harmon-Jones, 2006).

In sum, cognitive neuroscience research on control has highlighted different subcomponents of controlled processing. Although aspects of these subcomponents have been considered in previous social and personality psychology theories, the neuroscience literature has provided important clarifications of these components, as well as methods for assessing aspects of control that function rapidly and without conscious awareness.

Perception of the Self and Others

The Self: The Self is one of social-personality psychology’s earliest and most enduring constructs and, not surprisingly, it was among the first constructs to be examined in neuroscientific studies of social psychological processes (e.g., Craik et al., 1999; Klein, Loftus & Kihlstrom, 1996). Most of this research has examined brain activity associated with self-reflection and judgments about the self in comparison to judgments of others. Using PET, Craik et al. (1999) found that judgments of trait words as relating to the self vs. others were associated with large activations in regions of mPFC. In a similar study that used fMRI, Kelley et al. (2002) found that when comparing self-judgments of trait words with other-judgments (e.g., of George W. Bush), a region of ventral
mPFC activity was more strongly activated. The association between self-related processes and activity in ventral regions of the mPFC has since been replicated in several studies.

Research on processes related to the self—but not the self per se—such as agency, ownership, and valuation, have also revealed activation in the mPFC. This has led researchers to question whether the mPFC is involved merely in tracking self-concept (as research the role of mPFC in self-related judgments would suggest) or more broadly in some function or set of functions served by the self (Amodio & Frith, 2006). One study along these lines found that the act of choosing, whether for oneself or on behalf of someone else, invoked activity in mPFC, whereas having a choice made on one’s behalf did not (Nicolle, Klein-Flugge, Hunt, Vlaev, Dolan & Behrens, 2012). These and similar results have prompted the revised view that the role of the mPFC is to facilitate choice by tracking the value of possible outcomes (Ruff & Fehr, 2014), which is frequently confounded with “self” given that we usually make choices for ourselves. Thus, brain activations during self-related judgments may not reflect regions dedicated to the “self” per se, but rather regions that facilitate choice and behavior by tracking the value of choice-relevant attributes and features.

Perceiving faces. Information about conspecifics and social relationships is eminent in perception and cognition, and the initial stage of social processes often begins with face perception. Research on visual perception suggests that some components of the visual system are specialized for seeing faces, and that this component is localized to the fusiform gyrus in fMRI studies (Fig 4; Kanwisher, McDermot, & Chun, 1997). Although the idea of a specialized face area has been debated, with some arguing that fusiform responses to faces reflects expertise rather than a “face module” (Haxby, Hoffman, & Gobbini, 2000), the finding that this region responds to faces more than to other objects is consistent.
Faces are also known to elicit a characteristic ERP component that peaks 170 ms after the presentation of a face. This “N170” component is consistently larger to faces than non-face stimuli matched on other visual dimensions (Bentin, Allison, Puce, Perez, & McCarthy, 1996), making it a valuable neural marker of the engagement of low-level face-specific perception processes. The N170 is believed to reflect activity in multiple temporal-occipital structures linked to face processing, including the fusiform and other temporal regions (Fig. 1; Haxby et al., 2000; Puce, Allison, Gore, & McCarthy, 1995). Together, fMRI and ERP studies of face perception have shown that faces are perceived very quickly and at relatively low-levels in the hierarchy of visual processing.

Mentalizing and Theory of Mind. Once a person is encountered, the perceiver begins to evaluate and infer information about the individual. The process of inferring another person’s unique motives and perspectives is referred to as mentalizing, which is the process underlying one’s Theory of Mind (Frith & Frith, 1999). Early studies of mentalizing in the brain used PET to measure activity involved in inferences about the thoughts of other people or characters. These studies were motivated in part by an effort to understand autism and its links to deficits in Theory of Mind (Frith, 1989). In this line of research, Theory of Mind is best characterized by tasks involving false belief or deceptive intent – tasks on which successful performance depends on one’s ability to take another person’s perspective. In an early study on this topic, Fletcher et al. (1995) examined brain activity while normal subjects read a set of short stories. Mentalizing stories involved jokes or lies as a literary device – that is, they made sense to the extent that the reader understood that a character was the victim of a lie or joke. Hence, the stories required an understanding of a character’s false belief. Control stories did not rely on such devices, but rather involved straightforward physical descriptions. Although several brain regions were activated by both types of stories, only the mPFC was more active during the mentalizing stories. A similar set of
mentalizing activations were observed in another study when subjects viewed movies of people showing deceptive intent (Grezes et al., 2004).

Castelli, Happe, Frith, and Frith (2000) connected their findings on mentalizing with the attribution literature in social psychology. They measured brain activity while participants viewed a set of videos inspired by the famous Heider and Simmel (1944) animations, in which three shapes moved in an anthropomorphic fashion that implied human interactions. The authors found that the viewing of this type of animation also elicited mPFC activity, compared with control videos in which the movement of the shapes was not interpreted anthropomorphically. More recent work has further linked the process of dispositional attribution to activity in the mPFC (Harris, Todorov, & Fiske, 2005). Since these initial studies, a large body of research has associated activity of the mPFC with a range of tasks involving mentalizing and complex aspects of person perception and perspective taking (Amodio & Frith, 2006; Saxe, Carey, & Kanwisher, 2004). These tasks also typically elicit activity in regions of the superior temporal lobe (or temporal-parietal junction) and the temporal poles; these have been linked to the perception of biological motion and to conceptual representations of social information, respectively (Frith & Frith, 1999). Recent findings suggest that whereas the mPFC supports “internal” processes, such as mentalizing, these lateral regions (e.g., rTPJ) support “External” processes, such as perception and attention to social stimuli (Spunt, Falk, & Lieberman, 2010). Converging findings from the developmental literature corroborate the idea that mPFC development underlies the emergence of theory of mind abilities in children (Bunge et al., 2002), which in turn contribute to children’s social identity formation (Halim, Ruble, & Amodio, 2011).

Since the initial finding that mentalizing activated regions of mPFC, researchers have asked whether other forms of person perception might involve the same brain regions. A series of studies by Mitchell, Macrae, and colleagues proposed that social-cognitive aspects of person perception,
such as the ascription of trait attributes to a person, might also activate areas of mPFC (even if they do not necessarily require mentalizing). For example, when subjects judged noun-adjective word pairs that described a person, compared with those describing an inanimate object, activity was found in regions of interest within the mPFC, as well as areas of the temporal cortex and the temporal-parietal junction (Fig 3; Mitchell et al., 2002). Other researchers have observed activity in similar regions when simply viewing faces in an easy memory task (Gobbini et al., 2004), demonstrating that activity in this region in response to faces may not necessarily imply the inference of traits. Some research has found that viewing and making trait judgments of unfamiliar faces or dissimilar people is associated with activity in more dorsal regions of the mPFC, whereas more familiar and/or similar faces are associated with activity in more ventral regions (Gobbini et al., 2004; Mitchell et al., 2006). It is notable, however, that the mPFC is a large region of cortex, and the specific locus of person-related activity varies considerably across studies.

Reinforcement learning models of impression formation. Although much social psychological research on impression formation has focused on the passive learning of traits and behaviors, people often learn about others through the feedback they receive in direct social interactions. Recently, social neuroscientists have begun to study how we may form attitudes and trait impressions of people through mechanisms of feedback-based reinforcement learning (Hackel et al., 2015). For example, using a combination of fMRI and computational modeling methods, Hackel et al. (2015) found that, based on the positive or negative feedback received from another person in varying contexts, the value of the person (i.e., a basic attitude) as well as higher-level trait inferences are encoded in the ventral striatum, but that trait inferences additionally involve a broader network including right TPJ, precuneus, inferior parietal lobule, and ventrolateral PFC—regions previously identified as involved in more explicit forms of trait updating (Mende-Siedlecki et al., 2013). Moreover, when making decisions about future social interactions with these
individuals, participants’ choices were driven much more strongly by their trait impressions, and this choice process was underpinned by activity in the vmPFC, consistent with findings observed in nonsocial economic decision studies. The reinforcement learning approach has also been applied to understand how people use error signals to update their beliefs about the emotional states of others (Zaki et al., 2016). This new approach to understanding impression formation, which integrates social cognition, neuroscience, and computational models of reinforcement learning, promises to advance our understanding about how we learn about and act toward others in more realistic social interactions.

Empathy. Empathy is broadly defined as concern for another’s welfare (Batson, 1991), which may involve the process of experiencing another’s perspective and affective response (Lamm, Batson, & Decety, 2007). As with mentalizing and Theory of Mind, empathy is complex and involves a broad set of neural and psychological processes associated with affect, perception, social cognition, self-regulation, mimicry, and action (Decety, 2010). Many studies have examined empathy by measuring brain activity while a subject views another person experiencing pain. For example, Singer et al. (2004) used fMRI to measure brain activity while participants experienced a painful stimulus or viewed a loved one receiving the same stimulus. A set of structures, including the rostral ACC and anterior insula, were active in both conditions, relative to baseline, and activity in these areas to a loved one’s pain was greater among subjects with higher self-reported trait empathy. Other research suggests that similar brain regions were more active while watching racial ingroups experiencing pain than racial outgroups (Xu, Zuo, Wang, & Han, 2009). There are also suggestive findings from lesion patient studies, in which damage to the ventromedial PFC and ACC are associated with impaired empathy (Shamay-Tsoory, Tomer, Berger, & Aharon-Peretz, 2003). Given that the ACC is involved in a wide range of processes involving cognitive conflict and expectancy violation, these findings may reflect some aspect of expectancy violation or concern
when either the self or another person is subjected to pain, rather than suggesting that empathy is related to the experience of pain per se. Overall, this body of research has focused primarily on the brain mapping of empathic processes.

*Humanization.* Humanization refers to the process of seeing another person as possessing the characteristics unique to the human mind and the rights associated with being a member of society. Hence, dehumanization refers to the denial of such qualities to certain persons (Haslam, 2006). Members of one’s own social group are typically perceived as possessing these qualities, whereas members of a low-status outgroup are often seen as lacking many of these qualities (Leyens et al., 2001). The process of “humanization” is associated with empathy and mentalizing and typically refers to these processes as they relate to people. By comparison, empathy and mentalizing may also relate to non-humans and inanimate objects, as a form of anthropomorphism (Epley, Waytz, & Cacioppo, 2007). Research on dehumanization has connected the neuroscience work on mentalizing and the mPFC to important themes within social psychology concerning intergroup relations. For example, in a study by Harris and Fiske (2006), activity in a region of mPFC was greater when participants viewed pictures of valued others (e.g., members of the ingroup, people of high social status) compared with “dehumanized” individuals, such as drug abusers and homeless people.

*Section summary*

Social neuroscience research on social cognition and the self has begun to elucidate the rapidly-unfolding processes involved in perceiving and characterizing other people. By considering neural function and anatomy, and with the use of methods for assessing the very rapid processing of faces, this research shows that person perception often begins with the visual encoding of a face, often followed by affective responses, and that this information is drawn together with semantic knowledge about people and groups and the tracking of another’s mental state (e.g., in the TPJ) to
form trait representations in the mPFC. Representations of the self involve activity in ventral mPFC, a region that integrates and represents inputs from many brain regions to guide decisions and actions. By unpacking these different processes, this research helps to provide a more precise model of the core psychological components of social cognition and the self.

**Attitudes**

The attitude is a foundational construct in the field of social psychology, and attitudes were the subject of the earliest psychophysiological studies of social psychological processes. Rankin and Campbell (1955) measured changes in skin conductance levels (i.e., galvanic skin response) while White participants in their study interacted with White and Black experimenters. Skin conductance responses, which reflect activity of the sympathetic nervous system, were interpreted as states of heightened anxiety. Although participants in their study reported similar liking for the two experimenters, they exhibited larger skin conductance responses to Black than White experimenters on average. The authors interpreted the results as evidence for negative attitudes toward African Americans – the first evidence of implicit racial bias.

The attitude construct is complex, and although it is often defined simply as liking/disliking, multiple psychological and physiological processes may give rise to an attitude. The classic tripartite model of attitudes includes cognitive, affective, and behavioral components (Eagly & Chaiken, 1993). However, at the physiological level of analysis, an attitude likely corresponds to several other processes, such as motivational tendencies, action representations, changes in attention, and basic motor tendencies, among others. For this reason, it is difficult to map the high-level construct of an attitude onto specific physiological systems.

Early research on the neuroscience of attitudes used physiological measures to infer an individual’s affective or cognitive response to attitude objects. This approach allowed researchers to make unobtrusive assessments using a range of physiological indicators including skin conductance,
heart rate, facial EMG, EEG spectral power, and ERPs, among others (Cacioppo & Petty, 1983).

Much of this early work focused on psychological processes involved in persuasion, which leads to attitude change. Of special interest was the role of information processing. For example, Cacioppo and colleagues focused on asymmetries in parietal activity associated with semantic vs. non-semantic processing of a stimulus as an indicator of the degree of semantic processing of a message (e.g., Cacioppo, Petty, & Quintanar, 1982; in contrast to other researchers’ focus on hemispheric asymmetries in frontal cortical brain regions that were associated with emotion and motivation; e.g., Davidson & Fox, 1982).

In order to draw stronger inferences about the valence of a physiological response, facial EMG may be used to measure activity in muscle groups associated with different facial expressions linked to emotion (Dimberg, 1982; Cacioppo & Petty, 1979). Activity of the corrugator supercilii muscle, located just above and running parallel to the brow, is commonly described as brow-furrowing, and it is associated with many forms of negative affect, including anger and disgust. Activity of the zygomaticus major muscle, which runs across the cheek from the corners of the mouth to the cheekbones, is related to smiling, and it is associated with many forms of positive affect. Facial EMG has also been shown to be sensitive to micro-expressions – changes in facial muscle activity that are imperceptible to the eye (Cacioppo, Bush, & Tassinary, 1992). Thus, EMG provides information about an emotional expression, from which researchers may infer valence, approach/withdrawal orientation, or any other psychological process linked to a facial expression. Psychophysiological approaches such as these have been useful for assessing arousal and, to some extent, valence and cognitive processing, in order to infer attitudes. However, they were not designed to probe the specific neural mechanisms underlying attitude processes.

*EEG assessments of attitudes*
Early brain-based assessments of attitudes used measures of frontal cortical activity to assess affective responses to different stimuli and situations (Davidson & Fox, 1982). Responses to positive stimuli were associated with greater left-sided PFC activity, whereas responses to aversive stimuli were associated with greater right-sided PFC activity. Subsequent research by Harmon-Jones and his colleagues (Harmon-Jones, 2003; Harmon-Jones & Allen, 1998) suggested that, at this physiological level of analysis (i.e., left versus right PFC activity), many responses to positive and negative stimuli could be better understood in terms of approach-withdrawal processes rather than valence. This research raises questions about the meaning of the concept of “valence” at the neural or physiological levels of analysis.

ERP methods have also been used to examine attitudes. Early psychophysiology research observed that unexpected stimuli, such as a loud tone among a series of soft tones in an oddball task, typically elicit a positive-going ERP response approximately 300-500 ms after the deviant stimulus, maximal over parietal scalp sites (Squires, Donchin, Herning, & McCarthy, 1977). This ERP component was named the P300, for its positive polarity and its approximate peak at 300 ms. However, given that the latency of its peak varies considerably across tasks, it is often referred to as the P3 (i.e., the 3rd positive-going component following a stimulus onset) or late positive potential (LPP). The P3 was interpreted as reflecting the engagement of attention, associated with a surprise response, and the updating of one’s mental set (Donchin, 1981). More recent research has linked the P3 to activity of norepinephrine systems in response to an unexpected event (Nieuwenhuis, Aston-Jones, & Cohen, 2005), suggesting the P3/LPP reflects a complex set of processes associated with attention and arousal.

Cacioppo et al. (1993) noted that the LPP component, in conjunction with the oddball task, could be used to assess individuals’ evaluations of attitude objects. The authors found that a negative “oddball” stimulus that appeared within a series of positive objects elicited an LPP
response, compared with control stimuli. The LPP/oddball method of attitude assessment has even been shown to reveal people’s true attitudes when they attempted to misreport the attitude (Crites et al., 1995) or when implicit evaluations differed from explicit task instructions (Ito & Cacioppo, 2000). This technique of examining attitudes by measuring the LPP during the oddball task has since been used to assess a variety of attitudes, including intergroup attitudes (as described below).

**Neural substrates of attitudes**

More recently, fMRI research has probed the patterns of brain activity that underlie attitudes (Cunningham & Zelazo, 2007). Initial fMRI studies of attitudes sought to map the neural correlates of positive and negative word appraisals. For example, Cunningham et al. (2003) recorded brain activity while subjects viewed and evaluated pleasant or unpleasant words. They observed greater amygdala activity to negative than positive words. However, later findings suggested that it was the arousal associated with particular words that related to amygdala activity (cf. Whalen, 1998), whereas the valence of the words was associated with activity in the insula and frontal cortex, among other regions (Cunningham et al., 2004).

In more recent years, the field of neuroeconomics has greatly expanded our understanding of how people evaluate goods and money and how they process potential rewards. Although initially presented as research on economic decisions, rather than on attitudes, this field has become increasingly interested in the processes involved in social decision making (Ruff & Fehr, 2014). Indeed, by engaging subjects in economic games, often with real cash rewards, these studies offer a high degree of ecological validity regarding attitude formation and judgments, as well as greater insight on the operation of attitudes in social interactions.

Neuroeconomics research overwhelmingly implicates regions in the mesolimbic dopaminergic system, primarily the vmPFC, but also the orbitofrontal cortex (OFC) and ventral striatum (VS), in the integration of subjective value. (Indeed, the vmPFC and the OFC are
considered the same area by some, e.g., Pearson, Watson, & Platt, 2014.) Consistent with the common currency idea, this research suggests that the vmPFC is involved in the computation of subjective value of both appetitive and aversive stimuli (Bartra, McGuire, & Kable, 2013). In a series of studies, Rangel and colleagues have found that the vmPFC integrates information across a range of properties about a stimulus to produce a final value signal that integrates stimulus properties, active goals, costs, and other types of choice-relevant information (Rangel & Hare, 2010). Not only does vmPFC gather various value sources for a stimulus, but activity in the vmPFC also tracks the subjective value of a range of stimulus types. For example, vmPFC activity predicts choice regardless of whether the stimuli in question are food or money (Levy & Glimcher, 2011). More recently, the vmPFC was shown to integrate information about the reward value associated with a person and inferences about their traits to predict a perceiver’s desire for a future social interaction (Hackel et al., 2015). A related study found that activity in vmPFC scales with the subjective value of a monetary gain both for oneself and for another person (Zaki, Lopez, & Mitchell, 2014). These findings converge in identifying the vmPFC as playing a central role in the integration of subjective value.

**Emotion & Motivation**

Affect and emotion range from relatively low-level aspects of reward and punishment processes (e.g., approach vs. withdrawal states), to discrete basic emotions, such as joy, anger, and fear, to highly complex emotion processes such as guilt, jealousy, compassion, and schadenfreude. The evidence suggesting that sets of brain regions, hormones, and neurotransmitters may underlie discrete emotions is primarily based on non-human animal evidence (Panksepp & Biven, 2012). This type of research is more invasive and therefore more precise than the research that can be conducted with humans. There is evidence in humans regarding sets of brain regions being involved in discrete emotions that is obtained from functional magnetic resonance imaging (fMRI) of humans
lying supine while experiencing emotions (see Harmon-Jones & Peterson, 2009, for evidence showing how a supine posture reduces certain physiological responses associated with certain emotions). This fMRI evidence presents a more mixed picture as to whether there is any discrete emotion specificity with brain regions. That is, some meta-analyses have suggested there is no support (e.g., Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012), whereas other meta-analyses have suggested there is support for discrete emotions being linked to specific neural circuits (Vytal & Hamann, 2010). Hamann’s (2012, p. 458) interpretation of this literature is prescient, “… although neuroimaging studies have identified consistent neural correlates associated with basic emotions and other emotion models, they have ruled out simple one-to-one mappings between emotions and brain regions, pointing to the need for more complex, network-based representations of emotion.” Much like attitudes, emotion refers to complex psychological processes may not clearly map onto the organization of the brain and physiology. For this reason, most progress has been made in studying the neural processes associated with more basic forms of emotion and affect, including motivational, attentional, and autonomic processes, as well as lower-level processes associated with fear and reward.

Much affective neuroscience research is predicated on models that assume approach motivation and responses to rewards involve a positive affective system, whereas avoidance motivation and responses to punishments involve a negative affective system. The basic motivational dispositions toward approach and withdrawal are often associated with emotions. However, an emotion is not a “thing” but is a multi-component process made up of basic processes such as feelings of pleasure or displeasure, facial/body expressions, particular appraisals, and particular action plans and activation states (Frijda, 1993). Moreover, these components are not perfectly correlated with one another (Lang, 1995). For example, the same subjective feeling (e.g., anger) does not always imply the same behavior across contexts (e.g., at work versus on the soccer
field), and the bore of a neuroimaging scanner is a particularly limiting context in which to study the range of human emotions. In fact, as noted above, there is evidence that the supine position required for fMRI can mute emotional responding (Harmon-Jones & Peterson, 2009).

Approach and withdrawal motivational processes likely involve neural systems rather than specific brain structures. However, a systems-level analysis of approach-withdrawal motivation has yet to be thoroughly investigated due to the empirical difficulties of mapping these micro-processes in time. In addition, much of the research on the neuroscience of emotion and motivation has proceeded under a "brain mapping" approach rather than a "hypothesis-testing" approach as defined earlier. Consequently, this review will focus on brain regions that have received the most research attention. These are the amygdala, nucleus accumbens/ventral striatum, the orbitofrontal cortex, and asymmetrical frontal cortical regions. While considering the reviewed research, it is important to keep in mind the difficulty of making one-to-one associations between psychological and physiological processes. For example, if neurons in the amygdala become more active in response to viewing a face, it is almost impossible to claim that this activation reflects a certain psychological variable like fear, given that amygdala neurons become active in response to a wide range of psychological variables, including uncertainty, positive affect, and motivational relevance (Holland & Gallagher, 1999; Whalen, 1998). Perhaps even more importantly, the measurement of activity in neural regions are relatively imprecise in human neuroimaging and, consequently, important distinctions between various neurons and their psychological functions are almost impossible to study. For instance, the amygdala region can be divided into 13 nuclei and cortical areas (Swanson & Petrovich, 1998), and a basolateral region comprising the lateral, basal, and accessory basal nuclei has different, independent, and even converse activations than a central region comprising the central and medial nuclei (Fig 1; Moul, Killcross, & Dadds, 2012). Because of spatial
limitations of contemporary human neuroimaging (fMRI) equipment, researchers can only assess and interpret activity representing the entire amygdala region.

Perception of Motivational Relevance

Many of the stimuli that arouse motivation are perceived via the visual or auditory system. Novel and significant events attract our attention and engage the orienting response, and they are typically associated with approach and avoidance behavior (Thorndike, 1911) or the emotion evoked by an event (Bradley, 2009). The processes of orienting and attending have been posited to “stem from the activation of defensive and appetitive motivational systems that evolved to protect and sustain the life of the individual” (Bradley, 2009, p. 1). In this way, attention, emotion, and motivation are inextricably linked. Emotion is often theorized to be fundamentally a disposition to act, or to behave effectively to events that threaten or promote life (Frijda, 1986; Lang, 1985). The associated motivational tendencies are realized in general systems of approach and avoidance, with approach processes often acting to promote survival and avoidance processes often acting to prevent threats to well-being. Some theorists suggest that judgments of positivity reflect approach motivation, judgments of negativity reflect avoidance motivation, and judgments of arousal index the intensity of activation or motivation (Bradley, 2009). Although this may often be the case, the relationship between emotional valence and motivational direction (i.e., approach motivation is positive) is not always that direct. For instance, anger, a negatively-valenced emotion, is often associated with approach motivation (Carver & Harmon-Jones, 2009), a point to which we return later.

Fear processing

As noted above, human neuroimaging research has converged with the animal research to reveal that the amygdala is important for processing fear. For instance, the amygdala region is more activated by a neutral stimulus paired with an aversive event (conditioned stimulus) compared to
another neutral stimulus that does not predict an aversive event (LaBar et al., 1998). Moreover, 
amygdala activation correlates with the conditioned response of increased skin conductance (an 
indication of arousal) to the conditioned stimulus (LaBar et al., 1998). Going beyond these 
correlations, research has revealed that patients with lesions of the right, left, or bilateral amygdala 
do not demonstrate a conditioned response as measured by skin conductance even though they 
respond normally to the unconditioned (aversive) stimulus (Bechara et al., 1995). These results fit 
well with the animal research demonstrating that the amygdala plays an important role in fear 
conditioning.

Interestingly, although the amygdala is important for the acquisition of fear, as measured 
implicitly by skin conductance to fear-conditioned stimuli, it does not appear to be important for the 
acquisition of fear learning measured explicitly. Individuals who suffer bilateral amygdala damage 
acquire explicit knowledge about the relationship between the conditioned stimulus and the aversive 
unconditioned stimulus (Gazzaniga et al., 2002). This type of explicit knowledge is associated with 
the hippocampus (Squire & Zola-Morgan, 1991). Individuals with a damaged hippocampus but 
intact amygdala show normal skin conductance response to conditioned stimuli but no explicit 
knowledge of the relationship between the conditioned stimulus and unconditioned stimulus 
(Bechara et al., 1995).

Human neuroimaging research has revealed that the amygdala becomes activated in 
response to a variety of emotive stimuli in addition to fear-provoking ones (Holland & Gallagher, 
1999). For instance, experiments have revealed that positive stimuli also evoke greater amygdala 
activity than neutral stimuli (Breiter et al., 1996). Other studies have independently manipulated 
valence and intensity and found that amygdala activity is more associated with processing affective 
intensity than with processing any specific valence (Anderson et al., 2003). Consistent with results
obtained from these studies, Whalen (1998) observed that the amygdala is generally associated with vigilance to motivationally relevant stimuli.

Along these lines, research has revealed that extraversion is correlated positively with amygdala activation to positive (relative to negative) emotional pictures and that neuroticism is correlated positively with amygdala activation to negative (relative to positive) emotional pictures (Canli et al., 2001). Extraversion has also been found to correlate positively with amygdala activation to happy (relative to fearful) faces (Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002). Similarly, individual differences in promotion focus (sensitivity to gains) are correlated positively with amygdala activation to positive (relative to negative) word stimuli, whereas individual differences in prevention focus (sensitivity to losses) are correlated positively with amygdala activation to negative (relative to positive) word stimuli (Cunningham, Raye, & Johnson, 2005).

Again, it is important to keep in mind that the amygdala consists of multiple subnuclei that greatly differ in terms of structure, function, and connectivity with other regions (Davis & Whalen, 2001; McNaughton & Corr, 2004). Although these anatomical and functional distinctions may be observed using more invasive and precise methods in non-human animal research, the methods typically used in human research (e.g., fMRI, lesions) lacks the resolution to discern them. Even within the research literature on humans who have damaged amygdala, evidence is accumulating that damage to different sub-regions of the amygdala is associated with different psychological processes. For example, whereas past research on humans with damaged amygdala suggested that amygdala lesions impaired the perception of facial expressions of fear (Adolphs, 2008), recent research has revealed that individuals with selective damage to the basolateral amygdala are hypervigilant to facial expressions of fear and thus recognize fearful expressions better than do individuals without damage to the basolateral amygdala (Terburg et al., 2012). The previous research examined individuals with more extended or complete damage to the amygdala and thus
led to different – and diametrically opposite – conclusions than the more recent research that examined individuals with more selective damage to one region of the amygdala. Advances in both lesion research and in human neuroimaging will likely continue to reveal the diverse psychological functions of different sub-regions of the amygdala, consistent with research on rodents, which is more invasive and precise.

*Reward processing*

In humans, functional magnetic resonance imaging research has linked the anticipation of rewards to activity in mesolimbic dopamine regions, such as the nucleus accumbens and orbitofrontal cortex. For instance, Knutson, Wimmer, Kuhnen, and Winkielman (2008) found that anticipation of viewing rewarding stimuli (erotic pictures for heterosexual men) increased nucleus accumbens activity and financial risk taking. Nucleus accumbens activity also increased in anticipation of making a risky decision, that is, a high-risk ($1.00) as compared to a low-risk ($0.10) financial gamble. Moreover, the risk taking was partially mediated by increases in nucleus accumbens activation.

Research on dopamine and the nucleus accumbens have revealed differential mechanisms associated with the “wanting” and “liking” of a reward. Specific subregions of the nucleus accumbens in combination with specific neurotransmitters may be involved in “liking” or post-goal positive affect. For instance, microinjection of morphine, which activates opioid receptors, into posterior and medial regions of the accumbens shell increases positive affective reactions to sweet tastes (Peciña & Berridge, 2000). Other research has revealed that the nucleus accumbens is critical in regulating effort-related functions, such that lever pressing schedules that require minimal work are unaffected by accumbens dopamine depletions, whereas lever-pressing schedules that require high work are impaired by accumbens dopamine depletions (Salamone, 2007).
The orbital frontal cortex (Fig 2; OFC; often also called the vmPFC in decision-making and neuroeconomics studies) is another brain region that is widely implicated in studies of reward contingencies, such as in reversal learning, decision value, and in the self-monitoring of emotional responses. In reversal learning, an animal is taught that responding to one cue produces reward, whereas acting similarly to another cue produces nonreward or punishment. After the animal learns to respond correctly, the experimenter switches the cue–outcome associations, and the animal must learn to change its behavior. During cue-outcome learning across reversals, the OFC is activated (O’Doherty, Critchley, Deichmann, & Dolan, 2003).

In social psychological research, self-monitoring has been defined as the ability to evaluate one’s behavior in reference to higher order goals or other valued outcomes such as the reactions of other people (Snyder, 1974). This is the process “by which individuals evaluate their behavior in the moment to make sure that the behavior is consistent with how they want to behave and how other people expect them to behave” (Beer et al., 2006, p 872). Individuals with OFC damage have an impaired ability to prioritize solutions to interpersonal problems (Saver & Damasio, 1991), a tendency to greet strangers in an overly familiar manner (Rolls, Hornak, Wade, & McGrath, 1994), and to behave in disruptive manners in hospital settings (Blair & Cipolotti, 2000). They also tease strangers inappropriately and are more likely to include unnecessary personal information when answering questions (Beer, Heerey, Keltner, Scabini, & Knight, 2003). Moreover, people with OFC damage steeply discount delayed rewards (Sellitto, Ciaramelli, & di Pellegrino, 2010). The perspective that OFC is involved in bringing present-moment behavior in line with long-term standards and goals is consistent with the previously reviewed research on the emotional functions of this region.

*PFC asymmetries associated with emotion and motivation*
The asymmetric involvement of prefrontal cortical regions in positive affect (or approach motivation) and negative affect (or withdrawal motivation) was suggested many decades ago by observations of persons who had suffered damage to the right or left anterior cortex (Goldstein, 1939). Later research supported these observations using the Wada test, which involves injecting amytal, a barbiturate derivative, into one of the internal carotid arteries and suppressing the activity of one hemisphere. Amytal injections in the left side produced depressed affect, whereas injections in the right side produced euphoria (Terzian & Cecotto, 1959). These effects were interpreted as reflecting the release of one hemisphere from contralateral inhibitory influences. According to this view, activation in the right hemisphere, when not inhibited by the left hemisphere, caused depression; an uninhibited left hemisphere caused euphoria.

Subsequent studies appeared to confirm these results, finding that persons who had suffered left hemisphere damage or lesions tended to show depressive symptoms, whereas persons who had suffered right hemisphere lesions tended to show manic symptoms (Gainotti, 1972; Robinson & Price, 1982). Other research has revealed asymmetries underlying appetitive and avoidant behaviors in non-human animals, in species ranging from great apes (Hopkins, Bennett, Bales, Lee, & Ward, 1993) to spiders (Ades & Ramires, 2002).

More recent research suggests that in humans these affect-related asymmetric activations are often specific to the frontal cortex. This research often uses asymmetric activation in right versus left frontal cortical areas as a dependent variable, usually assessed by EEG recordings. Frontal cortical asymmetry is assessed by comparing activation levels between comparable areas on the left and right sides. Difference scores are widely used in this research, and their use is consistent with the amytal and lesion research described above that suggests that asymmetry may be the key variable, with one hemisphere inhibiting the opposite one.
Much of this evidence has been obtained with EEG measures of brain activity, or more specifically, alpha frequency band activity derived from the EEG. Research has revealed that alpha power is inversely related to regional brain activity using hemodynamic measures and behavioral tasks (Davidson, Chapman, Chapman, & Henriques, 1990). Source localization of EEG signals (Pizzagalli, Sherwood, Henriques, & Davidson, 2005) and fMRI results (Berkman & Lieberman, 2010) obtained in emotion-frontal asymmetry studies converge in suggesting that the dorsolateral PFC (Fig 3) is partially responsible for these effects. These findings are further corroborated by evidence from studies of transcranial magnetic stimulation, discussed later (Schutter, van Honk, d’Alfonso, Postma, & de Haan, 2001).

*Trait affective styles and resting frontal cortical asymmetry.* Depression has been found to relate to resting frontal asymmetric activity, with depressed individuals showing relatively less left than right frontal brain activity (Schaffer, Davidson, & Saron, 1983), even when in remission status (Henriques & Davidson, 1990). Other research revealed that trait positive affect in healthy subjects is associated with greater left than right frontal cortical activity, whereas trait negative affect is associated with greater right than left frontal activity (Tomarken, Davidson, Wheeler, & Doss, 1992).

Subsequent studies observed that trait approach motivation was related to greater left than right frontal activity at resting baseline (Amodio, Master, Yee, & Taylor, 2008; Harmon-Jones & Allen, 1997; Sutton & Davidson, 1997). These studies suggest that asymmetric frontal cortical activity could be associated with motivational direction instead of affective valence. However, avoidance and approach motivation are mostly associated with negative and positive affect, respectively (Carver & White, 1994), and consequently, the interpretation is clouded. Similarly, the finding that promotion (versus prevention) forms of regulatory focus are associated with greater relative left (versus right) frontal activation at baseline (Amodio, Shah, Sigelman, Brazy, &
Harmon-Jones, 2004) could be interpreted from either the motivational direction or affective valence view. That is, past research had essentially confounded emotional valence with motivational direction, but researchers had made the interpretation that relatively greater left than right frontal cortical activity reflected greater approach motivation and positive affect, whereas relatively greater right than left frontal cortical activity reflected greater withdrawal motivation and negative affect. These claims comport with dominant emotion theories that associated positive affect with approach motivation and negative affect with withdrawal motivation (Lang, 1995; Watson, 2000).

State affect and asymmetric frontal activity. Research has also demonstrated that asymmetric frontal brain activity is associated with state emotional responses. For instance, Davidson and Fox (1982) found that 10-month-old infants exhibited increased left frontal activation in response to a film clip of an actress generating a happy facial expression as compared to a sad facial expression. Frontal brain activity has been found to relate to facial expressions of positive and negative emotions, as well. For example, Ekman and Davidson (1993) found increased left frontal activation during voluntary facial expressions of smiles of enjoyment. Coan, Allen, and Harmon-Jones (2001) found that voluntary facial expressions of fear produced relatively less left frontal activity.

Some positive affects are lower in approach motivation, whereas others are higher in approach motivation. An important question remains regarding the findings relating affective valence to asymmetric frontal cortical activity: Do positive affects of any approach motivational intensity cause increases in relative left frontal activation? An experiment by Harmon-Jones, Harmon-Jones, Fearn, Sigelman, and Johnson (2008) addressed this issue by manipulating positive affect and approach motivation independently while measuring frontal EEG. Participants who wrote about positive/high-approach and positive/low-approach events both reported elevated positive affect, but only the positive/high-approach manipulation produced an increase in left-frontal EEG.
An experiment by Price, Hortensius, and Harmon-Jones (2013) found conceptually similar results but used facial expressions of high approach positive affect (i.e., determination; Harmon-Jones, Schmeichel, Mennitt, & Harmon-Jones, 2011) and low approach positive affect (e.g., satisfaction) to induce emotions. These results support the hypothesis that it is the approach motivational aspect of some forms of positive affect, and not the positive valence per se, that causes greater relative left frontal cortical activation (as measured by EEG).

**Anger and asymmetric frontal cortical activity.** The experiments described above suggest that approach motivation and positive affect are not perfectly associated with each other. More convincing evidence for the dissociation of affective valence and motivational direction (e.g., positive affect = approach motivation) comes from research on anger. Anger is a negatively-valenced emotion that evokes behavioral tendencies of approach (e.g., Ekman & Friesen, 1975). Anger is associated with attack, particularly offensive aggression (e.g., Berkowitz, 1993; Blanchard & Blanchard, 1984). Offensive aggression can be distinguished from defensive aggression, which is associated with fear. Other research also suggested that anger was associated with approach motivation (e.g., Amodio & Harmon-Jones, 2011; Izard, 1991). Furthermore, several studies have found that trait behavioral approach sensitivity (BAS), as assessed by Carver and White’s (1994) scale, is positively related to state and trait anger (Carver, 2004; Harmon-Jones, 2003;). Therefore, the emotion of anger provides a critical test case to disentangle interpretations of frontal cortical asymmetry as relating to valence or motivational direction.

To test this idea, Harmon-Jones and Allen (1998) assessed trait anger using a self-report questionnaire and assessed asymmetric frontal activity by examining baseline, resting EEG activity. In this study of adolescents, trait anger related to increased left frontal activity and decreased right frontal activity. Asymmetric activity in other regions did not relate with anger. The specificity of anger to frontal asymmetries and not other region asymmetries has been observed in all of the
reviewed studies on anger. These results have been replicated in other laboratories (e.g., Hewig, Hagemann, Seifert, Naumann, & Bartussek, 2004), and research has revealed that these results are not due to anger being evaluated as a positive feeling (Harmon-Jones, 2004). Other research has manipulated asymmetrical frontal cortical activity using transcranial magnetic stimulation and found that the disruption of right PFC increased approach responses to angry faces compared with disruption of the left PFC (van Honk & Schutter, 2006).

Researchers have also manipulated state anger to test the motivational direction model of frontal cortical asymmetry. Harmon-Jones and Sigelman (2001) found that individuals who were insulted evidenced greater relative left frontal activity than individuals who were not insulted. Additional analyses revealed that within the insult condition, reported anger and aggression were positively correlated with relative left frontal activity. Neither of these correlations was significant in the no-insult condition. Harmon-Jones, Peterson, and Harris (2009) conceptually replicated the above research and extended it by showing that social rejection causes increased relative left frontal activity that is associated with anger and jealousy.

Considered as a whole, EEG asymmetry studies of emotion and motivation have provided a critical test of how emotions are organized in the brain. Although the subjective appraisal of emotion is usually described in terms of valence (e.g., pleasant vs. unpleasant), this large body of research demonstrates that at the neural and physiological level of analysis, emotional responses are organized in terms of their approach vs. withdrawal motivational orientation.

*Emotion regulation*

The idea that people often seek to regulate their emotional response dates back to classic theories of Freud and Descartes, who suggested that reason is needed to overcome passion (in Freud’s terms, the Ego must adjudicate the sway of the Id). Contemporary views on emotion regulation have similarly focused on the role of cognitive control in modulating lower-level
emotional processes (Davidson, Jackson, & Kalin, 2000). In social cognitive neuroscience, researchers have focused on regions of the PFC as the neural substrate of control and on regions of the subcortex, such as the amygdala, as the substrate of emotion, particularly negative emotion (Ochsner & Gross, 2008). For example, in many emotion regulation studies, subjects view aversive (vs. neutral) images during a functional magnetic resonance imaging (fMRI) scan. On some trials, subjects are instructed to simply view the image. On other trials, subjects may be instructed to decrease their affective response to the image.

There is now ample evidence of the general pattern that activation in lateral prefrontal cortical regions is inversely related to activation in subcortical regions during emotion regulation attempts, and that the degree of inverse association is linked to emotion regulation success (see Buhle et al., 2014, for a meta-analysis). For example, emotion regulation using cognitive reappraisal of upsetting negative images (Ochsner & Gross, 2008) and appetitive food cues (Giuliani, Mann, Tomiyama, & Berkman, 2014) increases activity in dorsolateral and ventrolateral prefrontal cortices and decreases activity in amygdala and ventral striatum, respectively, and increases the inverse coupling between the two systems (Banks, Eddy, Angstad, Natha, & Phan, 2007). When measured, activity in lateral prefrontal regions and its connections with subcortical regions also tracks with self-rated control success (Ochsner et al., 2004). Similar patterns conforming to the opposition assumption have been observed among cigarette smokers controlling urges to smoke (Kober, Mende-Siedlecki, Kross, Weber, Mischel, Hart, et al., 2010), cocaine abusers controlling drug craving (Volkow, Fowler, Wang, Telang, Logan, Jayne, et al., 2010) and in clinical populations (e.g., Goldin, Manber-Ball, Werner, Heimberg, & Gross, 2009). At a first pass, this broad pattern is consistent with the theory that lateral prefrontal regions down-regulate subcortical structures for both appetitive (e.g., cravings) and aversive (e.g., distress) responses.
However, despite observed correlations between regions of lateral PFC and the amygdala, these regions share few direct anatomical connections, and of these sparse connections, most run from the amygdala to the PFC (Ghashghaei & Barbas, 2002). Instead, “down-regulation” of the amygdala and other subcortical structures such as the ventral striatum may occur through indirect routes. For example, self-regulation of fear by lateral prefrontal regions is to be mediated by vmPFC, apparently “piggybacking” on an evolutionarily primitive fear extinction pathway (Delgado, Nearing, LeDoux, & Phelps, 2008). Other studies of emotion regulation have found similar indirect pathways of lateral prefrontal effects on subcortical regions being mediated through more ventral regions involved in valuation (Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008). Still other theoretical analyses suggest that PFC processes function primarily to modulate attention and action in response to an emotional event, rather than directly suppressing the internal affective signal from subcortical structures (e.g., Amodio & Ratner, 2013), a perspective consistent with previous physiological studies (Gross & Levenson, 1993).

The neuroscience research on emotion regulation highlights the idea that psychological conceptions of an emotion may not map cleanly onto any specific underlying neural structure (Amodio, 2008; Cacioppo et al., 2003). As we noted above, the constructs of emotion and emotion regulation, as they are understood by psychologists, do not appear to represent mechanisms that can be mapped to unique neural processes. Rather, emotions and emotion regulation may reflect a summary appraisal of several interacting physiological mechanisms that function in concert to promote adaptive responses to an emotion-eliciting situation. If so, then questions about the neural correlates of emotion and emotion regulation may need to be reconsidered and perhaps replaced by questions about more basic processes of motivation, attention, arousal, and action that accompany subjective experiences of emotion.

**Intergroup processes**
One of the most active areas in the field of social neuroscience examines prejudice, stereotyping, and intergroup relations (Amodio, 2014). The area has provided fertile ground for social neuroscience research because it concerns the confluence of multiple psychological processes across multiple levels of analysis. That is, researchers can simultaneously examine issues of automaticity and control, emotion and cognition, motivation, attitudes, and a range of individual differences, all within the context of social cognition and social behavior, and with broad societal implications.

*Seeing groups*

Social interactions often begin with the perception of a face. Indeed, mounting evidence reveals that social motivations can alter the way a face is seen (Freeman & Johnson, 2016; Hugenberg et al., 2008; Kawakami, Amodio, & Hugenberg, in press), an effect that likely reflects the modulatory influences of signals from the temporal cortex and prefrontal cortex to the fusiform gyrus (Bar et al., 2006). This insight suggests that group memberships, and perhaps also the prejudices and stereotypes linked to groups, may alter early-stage visual face processing.

Early functional MRI (fMRI) research demonstrated greater fusiform activity in response to faces of one’s own racial group (i.e. the ingroup) — an effect that was associated with better recognition of ingroup than outgroup faces (Golby et al., 2000). Research examining the N170 component of the event-related potential (ERP), which indexes the degree of initial configural face encoding at just ~170 ms, revealed enhanced processing of ingroup vs outgroup faces, even when groups were defined arbitrarily (Ratner & Amodio, 2013). This finding was consistent with fMRI data showing that faces of ‘coalition members’ elicited greater activity in the fusiform gyrus than did other faces, regardless of race (Van Bavel, Packer, & Cunningham, 2011). Hence, social group membership, even when defined on the basis of minimal categories, appears to promote greater visual encoding of a face. These findings dovetail with behavioural research showing that biased
visual representations of outgroup members facilitate discriminatory actions toward them (Krosch & Amodio, 2014; Ratner et al., 2014).

Group effects on face processing become more complex with existing relations. Whereas outgroup faces in minimal group contexts elicit reduced processing, faces of potentially threatening outgroup members may elicit greater processing. In the context of race, racial outgroup members are often viewed as threatening, and therefore often elicit vigilant attention (Trawalter, Todd, & Richeson, 2008). This effect is also reflected in perceptual encoding of a face. For example, larger N170 ERP amplitudes to viewing black vs white faces (equated in luminance) have been observed in subjects with stronger implicit prejudice (Ofan, Rubin, & Amodio, 2011) and in subjects who were made to feel anxious about appearing biased (Ofan, Rubin, & Amodio, 2014). These and other findings suggest the visual processing of race is malleable and depends on social motivations and context (e.g., Senholzi & Ito, 2013; Schmid & Amodio, 2016; Walker et al., 2008).

Neural representations of race (e.g., Black vs. White), determined with multi-voxel pattern analysis (MVPA), have been observed in the fusiform gyrus, and these neural representations have been associated with behavioural indices of implicit prejudice and stereotyping (Gilbert et al., 2012). Together with other recent MVPA studies of race perception (e.g., Contreras et al., 2013; Van Bavel et al., 2013), this research further demonstrates that group membership appears to influence the earliest stages of face processing.

The finding that social category cues modulate the early visual processing of ingroup and outgroup members’ faces becomes especially important given that this effect occurs spontaneously and unconsciously, and thus any biases in the initial perception of a face may be assumed by the perceiver to be veridical. This suggests a new route through which implicit racial biases may affect behaviour, and it reveals a new challenge for methods of prejudice control that require that bias be detected before regulatory processes are engaged.
Neural basis of implicit prejudice

Physiological research on responses to race provided the first evidence for implicit forms of racial bias that differed from explicit self-reports (Rankin & Campbell, 1955; Vanman, Paul, Ito, & Miller, 1997). The earliest neuroscience studies of intergroup bias attempted to map this implicit response to the brain. At the time of these early neuroimaging studies, new findings were emerging from the study of fear conditioning in non-human animals that identified the amygdala as a key neural structure involved in the fear response (Davis & Whalen, 2001; LeDoux, 2000). For these reasons, the amygdala was initially considered to be a likely neural substrate of prejudice. However, although early neuroimaging studies of prejudice were suggestive of the amygdala’s role in implicit bias, evidence for this connection has been mixed and often unreliable. Indeed, we now know that the neural basis of prejudice involves a more complex network of neural processes whose operation depends on the social context and an individual’s goals (Amodio, 2014).

The possibility that the amygdala underpinned implicit prejudice was initially tested in a pair of studies, by Hart et al. (2000) and Phelps et al. (2000), which used fMRI to measure participants’ brain activity while viewing faces of ingroup and outgroup members. Interestingly, neither study found direct evidence for differential amygdala activity in response to ingroup and outgroup members. Hart et al. (2000) found that while ingroup and outgroup faces both elicited a similar increase in amygdala activity compared with baseline, this level of activity habituated more quickly in response to ingroup faces than outgroup faces—indirect evidence for greater amygdala activity to racial outgroups. In Phelps et al. (2000), amygdala activity did not vary as a function of a target person’s race, again suggesting no direct role of amygdala activity. However, when the authors examined the subset of amygdala voxels that increased in activity to Black compared with White faces for each participant, this selective activity correlated with participants’ scores on an IAT measure of implicit prejudice, as well as with their degree of startle-eyeblink modulation to Black
vs. White faces. None of these measures were associated with explicit self-report measures of racial attitudes, suggesting that these patterns of amygdala activation may represent an implicit association. That said, these seminal studies were conducted in the early days of fMRI, with very small samples, using analytical methods that have since been improved.

From a social-personality perspective, identifying a neural substrate of implicit racial bias was interesting because it might offer new clues as to how implicit racial biases are learned, activated, expressed in behavior, and potentially altered or extinguished. Furthermore, such research could help to clarify important individual differences in people’s abilities to respond without prejudice. For example, it could clarify whether egalitarians respond without prejudice because they are very effective in regulating their responses or because they do not show such responses in the first place.

To begin to address such questions, Amodio, Harmon-Jones, and Devine (2003) examined implicit responses to racial ingroups vs. outgroups using a startle-eyeblink index of amygdala activity, rather than fMRI. The startle eyeblink—a component of the whole-body startle reflex that is amplified when a person is in a threatened state—is mediated by a neural circuit that specifically involves the amygdala’s central nucleus, the subregion most directly involved in the fear response (Davis, 1992). By contrast, fMRI is currently unable to discern the amygdala subnuclei specifically involved in threat processing.

Using the startle eyeblink method, Amodio et al. (2003) observed significantly stronger responses to Black compared with White or Asian faces among White subjects, providing evidence that the amygdala may underlie implicit fear responses to Black outgroup members. This finding also helped to establish the existence of an affective form of implicit bias that could be discerned from semantic forms of evaluation and stereotyping, as well as from an individual’s explicit racial attitudes (Amodio & Devine, 2006; Fazio et al., 1995; Greenwald & Banaji, 1995). Because the
amygdala response, as revealed using startle eyeblink, represents an implicit process, this finding also helped to explain how people can learn and express implicit prejudices that may conflict with their explicit beliefs.

Since these initial studies, several event-related fMRI studies have observed that amygdala activity is greater to outgroup than ingroup faces under some conditions but not others, and that these effects are typically subtle. That is, differences in amygdala activity to Black vs. White faces tend to emerge only in the context of a minimally demanding task. As tasks become more demanding, such as when searching for a “dot” on the image, when imagining whether the target likes a particular vegetable, or when attempting to match the face to written group labels, amygdala activations tend not to be found (Cunningham et al., 2004; Lieberman et al., 2005; Wheeler & Fiske, 2006). Other research suggests that amygdala effects are lessened when the outgroup face is not looking directly at the subject (i.e., has averted gaze; Richeson, Todd, Trawalter, & Baird, 2008). And, several studies have found no differences in amygdala responses to racial ingroup vs. outgroup faces (e.g., Gilbert et al., 2012; Richeson, 2003; Ronquillo et al., 2007).

The mixed findings on the role of the amygdala in implicit prejudice suggests that this line of inquiry remains a work in progress. One plausible alternative explanation is that amygdala activity reflects not a fear of an outgroup face, but rather a participants’ concern about appearing prejudiced in the eyes of themselves or others. Other research has shown that the amygdala responds to motivationally-relevant stimuli, such as to rewarding stimuli (Holland & Gallagher, 1999) or to one’s ingroup members (e.g., Van Bavel et al., 2008), although it is unclear whether these responses involved the basal nucleus, which coordinates attention for approach behavior, rather than the central nucleus of the amygdala, which coordinates a threat response. More broadly, these findings highlight the complexity of neural and psychological processes and the need for careful interpretations of neuroimaging data (Amodio & Lieberman, 2009).
As this literature develops, researchers have begun to explore the role of other neural regions such as the mPFC, related to mentalizing and humanization, the orbital frontal cortex, linked to evaluation and decision making, the insula, associated with feelings such as disgust, and the striatum, a structure associated with representations of value and appetitive behaviors (see Amodio, 2014, for a review). These emerging findings reveal a network of structures involved in prejudiced attitudes, feelings, and judgments. As this work continues, we hope that knowledge about the neural substrates of prejudice will shed light on how prejudices are formed expressed, and reduced.

Stereotyping. Whereas much research has examined the neural correlates of race-related affect and evaluation, relatively fewer have investigated social stereotypes (Quadflieg & Macrae, 2011). Stereotypes are cognitive structures stored in memory that represent a set of attributes associated with a social group (Hamilton, 1981). Amodio and Devine (2008; see also Amodio, 2008; Amodio & Mendoza, 2010) proposed that stereotypes are rooted in mechanisms of semantic memory and selection, which are associated with neural activity in the temporal lobes and lateral PFC, respectively. Behavioral and neuroscience research on semantic learning systems has uncovered the dynamics of how such associations are learned and expressed in behavior. By linking stereotypes to this literature, researchers can apply findings from the memory literature to understand stereotyping processes (Amodio & Ratner, 2011a). For example, whereas affective associations are learned quickly and are relatively indelible, semantic associations may be learned and unlearned through a process of repeated pairings and non-pairings. Semantic learning systems are more likely to be expressed in trait impressions, goal representations, and goal-driven behaviors, and thus they are more likely to emerge in verbal responses (Amodio & Devine, 2006).

More recently, fMRI studies of social stereotypes are beginning to illuminate the major neural substrates key neural regions. The anterior temporal lobe (i.e., the temporal poles) have been found to represent knowledge about people and social groups (Olson et al., 2013). The dorsal part of
the ATL, which is implicated more specifically in the representation of social objects (i.e., people), is densely interconnected with the regions of the mPFC that are associated with trait judgment and impression formation (de Schotten, Dell’Acqua, Valabregue, & Catani, 2012). This suggests that social information represented in the ATL is selected into the mPFC to support the process of social cognition.

Not surprisingly, the ATL is consistently implicated in studies of stereotype representation. In one fMRI study (Gilbert et al., 2012), the authors used multivoxel pattern analysis (MVPA) to examine neural activity representing judgments of black and white individuals on the basis of stereotypic traits (athleticism) vs evaluations (potential for friendship). Results showed that both forms of person judgment were represented in the left ATL, and that these neural representations correlated with behavioural measures of implicit racial stereotypes and implicit racial attitudes, respectively. In a different study (Contreras, Banaji, & Mitchell, 2012), participants considered either social or non-social categories (e.g., men vs women; violins vs guitars) and judged which category was more likely to be characterised by a particular feature (e.g., enjoys romantic comedies; has six strings). A region of ATL was more strongly activated during stereotype-relevant judgments of social categories. Consistent with these findings, the disruption of ATL activity by transcranial magnetic stimulation attenuated the behavioural expression of implicit gender stereotype associations (Gallate et al.), suggesting that the ATL is necessary for stereotype representation. Thus, knowledge of social stereotypes appears to reside in the ATL.

Other research has begun to explore the roles of additional neural structures, including the mPFC, a region associated with mentalizing and trait impression formation. Through its connections with the ATL, the mPFC may support the activation of stored stereotypic knowledge into a working mental representation during person perception. In addition, regions of the lateral PFC linked to the selection of semantic information into working memory may also support the
activation and accessibility of stereotype knowledge. Researchers are just beginning to understand the neural substrates of stereotyping, but it is already clear that stereotyping is supported by a network of neural regions rather than a single structure, suggesting a complex and multifaceted process.

Control and the Regulation of intergroup bias

Given that racial stereotypes and implicit evaluations may be automatically activated, regulatory processes are needed to reduce their expression in behavior. Social-personality psychologists have begun to apply findings from neuroscience studies of cognitive control to understand how expressions of intergroup bias may be regulated (Amodio, Devine, & Harmon-Jones, 2007; Amodio & Devine, 2010). The hope is that a better understanding of the mechanisms of control will inform efforts to reduce prejudice.

Detecting the need for control. Neuroscience theories of control maintain that it involves two general components: the detection of conflict and the implementation of an intended response (Botvinick et al., 2001). This neuroscience model was initially applied to the issue of prejudice control by Amodio et al. (2004), who noted that expressions of bias could reflect a failure to detect the need for control by the conflict monitoring system or rather a failure to implement control once its need has been detected. Depending on the underlying cause, a different kind of intervention may be needed (Amodio & Devine, 2005).

Amodio et al. (2004) tested this model of prejudice control by measuring EEG among low-prejudice participants while they completed the weapons identification task – a sequential priming task that requires enhanced control on some trials to override an automatic stereotyping bias. The use of a stereotype control task was critical because it (a) clearly manipulated the engagement of controlled processing and (b) provided separate behavioral indicators of automatic stereotyping biases and the degree of successful response control – both of which are needed to interpret
associated brain activity as being involved in control. Without such a task, inferences cannot be made about the role of brain activity in control. By examining event-related EEG signals, they found that the dorsal ACC was more strongly activated on trials where control over stereotypes was needed. Importantly, heightened ACC activity was observed both when control succeeded and when it failed. In the case of failure, the finding suggested that when low-prejudice people respond with unwanted stereotypes, their brain may register response conflict and the need for control, but efforts to exert top-down control are impaired. A conceptually related ERP study by Bartholow, Dickter, and Sestir (2006) found that alcohol administration selectively impaired the regulative component of control without affecting the conflict monitoring component in the process of stereotyping inhibition, providing further support for the distinction between these two components of control in the regulation of intergroup bias.

In subsequent research, this model has been used to address why some low-prejudice people – those who have strong internal motivation to respond without prejudice but are also very concerned about external social pressures – are especially prone to unintended expressions of bias (Amodio et al., 2008), and also why it is more difficult to detect the need for control when one regulates prejudices based on external social cues (Amodio, Kubota, Harmon-Jones, & Devine, 2006). Finally, several other ERP and fMRI studies have observed ACC-related activity linked to the perception of race (e.g., Cunningham et al., 2004; Lieberman et al., 2005; Richeson et al., 2003), and although these studies did not directly assess controlled processing, the pattern of results is consistent with the conflict monitoring model.

Implementing control. Once the need for control is detected, other mechanisms are engaged to implement a controlled response (Devine, 1989; Shiffrin & Schneider, 1977). Neuroscience research has implicated the PFC in this function (Botvinick et al., 2001; Kerns et al., 2004; Badre & Wagner, 2008). However, the specific target of control is not always clear – it could be the
emotional response, the stereotype itself, the expressed behavior, the way a person is perceived, or some other process. An analysis of the neuronal circuitry of the PFC provides clues about the targets of control. In particular, this circuitry suggests that the PFC modulates goal-directed action processes as well as the modulation of sensory input and perceptual processing (Miller & Cohen, 2001). By contrast, the PFC has few connections to the amygdala, the region often identified as a neural correlate of implicit bias. Thus, neuroanatomy suggests that the regulation of prejudice involves the modulation of perception, attention, goals, and actions.

Several studies have observed patterns of PFC activity associated with responses to race, but their role in response control has not been clear. In the earliest example, Richeson et al. (2003) found that subjects who responded with stronger PFC activity when viewing Black vs. White faces in one experimental session were more likely to perform more poorly on a cognitive control task following a stressful interracial interaction in a different experimental session. The authors reasoned that subjects who spontaneously engaged control when viewing faces were also more likely to engage control during an interracial interaction, which in turn interfered with their performance on the Stroop task. In a different study, neural responses to faces of Black and White people’s faces were measured using fMRI (Cunningham et al., 2004). The authors observed greater activity in the ACC and regions of PFC to Black faces compared with White faces, suggestive of the engagement of control.

Other research has examined the role of the PFC in modulating behavioral intentions and the perceptual processing of race in a way that is more consistent with PFC anatomy. Amodio et al. (2007) demonstrated that increased left PFC activity was associated with the behavioral intention to engage in prejudice-reducing behaviors. More recently, Amodio (2010b) proposed that mechanisms of control promote intentional behavior by modulating attention to and perception of cues that control is needed. For example, Monteith’s (1993) self-regulation model posited that, once the goal
to control intergroup responses is formed, an individual becomes vigilant to cues that control is needed, such as the presence of an outgroup member. Amodio (2010) predicted that control-related PFC activity would serve to allocate greater attentional resources to the perception of outgroup faces, which in turn would facilitate better response control (i.e., more accurate responding despite any biasing effects of racial stereotypes). Indeed, while participants completed the weapons identification task, greater left PFC activity throughout the task predicted larger attentional ERP responses to Black vs. White face primes just 180 ms after a face appeared, as well as greater response control.

In summary, social neuroscience research on the regulation of intergroup bias is notable for providing some important advances to social-personality theory; for example, it has introduced the conflict monitoring model of control to the regulation of social behavior and has used this approach to explain puzzling individual differences in people’s ability to inhibit unwanted biases.

**Personality and Individual Differences**

Personality is the study of enduring psychological dispositions and their influence on thought, emotion, and behavior. For researchers interested in personality processes, physiological processes that operate on a longer timecourse, such as genes and hormones, are of special interest. Some research in this area has examined direct correlations between personality traits, such as those comprised by the “Big Five.” Increasingly, however, researchers have used neural and physiological models of continuity and change to help understand the complexities of personality processes and individual differences. In this section, we describe just a few examples of this growing area of research.

**Affective style**

Research by Davidson and colleagues suggested that enduring affective styles, associated broadly with depression and anxiety, related to different patterns of neural function (Davidson &
Irwin, 1999). Specifically, these styles relate to differences in frontal cortical asymmetries, as described in previous sections, which are also associated with differential approach vs. withdrawal tendencies (Davidson, 1998). Although much research has examined state changes in frontal EEG asymmetry to study emotion and motivation, more stable, trait-like components of the asymmetry have been taken to resemble a substrate of affective personality style. These trait-like patterns have been observed in adults, children, and non-human primates using a range of measures, and individual differences in these styles have been linked to various assessments of mental and biological health (Kern et al., 2008). More recently, trait-like patterns of activity in other brain regions, including the amygdala and regions of the basal ganglia have been included in an expanded framework of affective style (e.g., Fox et al., 2008).

Role of genes in personality and social psychology

Psychologists have long suspected that many personality traits are substantially heritable, with longitudinal studies showing strong continuity in temperament from childhood to adulthood (Cramer & Block, 1998). Over the past several decades, research on heritability using twin designs has supported this view (Caspi, Roberts, & Shiner, 2005; Plomin, DeFries, McClearn, & Rutter, 1997). Across the “Big Five” personality dimensions, a review of heritability estimates suggests that these traits are approximately 50% due to genetic similarities (Bouchard & Loehlin, 2001). The traits neuroticism and extraversion tend to show the largest heritability estimates, consistent with theory and research suggesting that these factors are dominant over other traits (Eaves, Eysenck, Martin, 1989; Tellegen et al., 1988). Significant heritability has also been observed for attitudes, such as toward social policy (e.g., the death penalty, immigration), racial beliefs, and hobbies (e.g., doing crossword puzzles) (Olson, Vernon, Harris, & Lang, 2001).

More recently, research has begun to explore particular gene polymorphisms found in DNA to examine their correlation with personality traits (Plomin & Caspi, 1999). The most well-known
example of this candidate-gene approach is the effort to associate differences in emotional processing with variations in the serotonin transporter gene – a gene that codes for proteins involved in the reuptake of serotonin from the synapse (Hariri & Holmes, 2006; Canli & Lesch, 2007). Lesch et al. (1996) found that trait neuroticism was associated with individual differences (i.e., polymorphisms) in this gene. Other research has observed suggestive associations between gene polymorphisms related to dopamine function and traits of extraversion and sensation seeking (Smillie et al., 2010), although a meta-analysis suggests that evidence for these relationships is mixed across studies (Munafò et al., 2008). Another method for garnering convergent insight into the genetic and neurotransmitter systems involved in social processes is the experimental administration of various drug challenges. These studies provide insight into the molecular involvement in aspects of personality and various behaviors, such as extraversion and economic decision making (DePue et al., 1994; Crockett et al., 2008).

A growing body of literature suggests that environmental experiences can directly and indirectly modulate the expression of DNA (Caspi et al., 2003; Champagne & Curley, 2005), consistent with psychological theories that highlight the importance of both person and environmental factors. Although this approach has generated much excitement, it is notable that effects in this literature have been difficult to replicate (e.g., Caspi et al., 2005). It is likely that the extreme complexity involved in traversing such distal levels of analysis – from DNA to complex behaviors, traits, and mental states – remains far beyond the grasp of extant theoretical models. Genes merely code for proteins, after all, not behaviors. Therefore, much of the current work in this area continues to explore gene-behavior associations in an effort to incrementally constrain our understanding of their causal relationship, slowly progressing toward a coherent genetic account of personality and behavior.

Hormones and psychological dispositions
Although genes provide a close analog to the concept of personality as an enduring trait, the role of hormones in individual differences and social behavior has received much more attention. Hormones are characterized as providing a broader “organizational” function, in that they help to orchestrate the coordinated response of multiple physiological and brain mechanisms. By comparison, specific neural activations are typically interpreted as reflecting very specific, low-level aspects of a psychological response. Furthermore, whereas neural processes typically relate to specific state-related responses, the effects of hormones on behavior are slower, ranging from a few seconds, in the case of hormonal responses to specific events, to the course of a lifetime, in the case of baseline hormonal function. Hormonal influences in early development can set the stage for enduring dispositions in biological and mental processes. For example, prenatal exposure to sex hormones has been shown to have long-term implications for gender development and adult sexual behavior (Singh, Vidaurre, Zambarano, & Dabbs, 1999).

As with gene effects, hormone effects can vary substantially as a function of the situation. For example, testosterone levels may vary with changes in power and social status (Mazur & Booth, 1998), and individual differences in the testosterone response following a competition predict whether they choose to seek a rematch (Mehta & Josephs, 2006). Experimental administration of testosterone can also increase attention to potential social threats, such as angry faces (van Honk et al., 1999). More recent research suggests that the hormones testosterone and oxytocin may play a larger roles in orchestrating social behavior, such that they may promote greater cooperation in reciprocal social exchanges (de Dreu et al., 2010).

Another steroidal hormone, cortisol, is widely studied as a physiological response to stress (Dickerson & Kemeny, 2004). Cortisol is produced by the adrenal glands following activation along the hypothalamic-pituitary-adrenal axis, and it functions broadly to regulate metabolism and arousal in dispositional (i.e., baseline) diurnal processes and also in response to specific arousing events.
Cortisol secreted into the blood can be detected in saliva, and thus salivary cortisol concentrations may be measured non-invasively and with relatively low cost in the typical psychological laboratory (Schulthiess & Stanton, 2009).

Cortisol reactivity in response to a stressor coordinates an adaptive response (e.g., fight or flight), but after chronic exposure this response becomes maladaptive (McEwen, 1998). For example, higher baseline cortisol levels have been associated with unhealthy profiles, including perceived stress, anxiety, depression, and cardiovascular stress (Cohen et al., 2006). Although cortisol reactivity in response to a stressful event has been examined in different contexts, a meta-analysis by Dickerson and Kemeny (2004) suggests that it is especially sensitive to socio-evaluative stressors, such as when a subject must give an extemporaneous speech to a panel of disapproving peer judges. Thus, cortisol research has highlighted the primacy of social interactions in human motivation and stress responses. Also, there is evidence that cortisol interacts with testosterone in the context of competition and status-seeking (Mehta & Prasad, 2015), suggesting that the stress system might modulate the effects of other important biological regulatory systems. As an outcome measure, cortisol provides a useful assessment of the stress response that does not rely on self-report. Furthermore, the connection between psychological distress and biological response highlights connections between the mind and body, and it underlines the important effects that social and dispositional factors have on physical health.

Research on immune variables, such as proinflammatory cytokines, provides a similar link that is more specific to healing and illness processes (Maier & Watkins, 1998). Ratner, Halim, and Amodio (2013) examined hormonal and immune responses in African American and Latina women to investigate the impact of perceived stigmatization on health. They found that reports of greater stigmatization were associated with higher baseline levels of IL-6, a pro-inflammatory cytokine that, when chronically elevated, increases vulnerability to disease. At the same time, reports of pro-
ingroup attitudes were associated with elevated DHEA, a hormone associated with cell repair and resilience. Using this approach, Ratner et al. (2013) revealed a harmful effect of social stigmatization that could be buffered by one’s social identity. Other research has shown that individual differences in emotion regulation strategies relate to health risk via associations with cytokine activity (Master, Amodio, Yee, Stanton, Hilmert, & Taylor, 2009).

It is notable that the greatest power of the social neuroscience approach lies in its ability to probe mechanism, and researchers have recently begun to conceptualize hormones and immune variables as mechanism variables. For example, Amodio (2009) measured salivary cortisol as a downstream correlate of within-brain norepinephrine activity to test a model of how intergroup anxiety might affect norepinephrine-related mechanisms of control involved in the inhibition of stereotypes. Maier and Watkins (1998) provided a detailed analysis of how changes in cytokines and other associated immune and endocrine variables can act as mechanisms to alter cognition, emotion, and behavior, and to promote an organism’s health. These examples suggest the potential of endocrine and immune approaches to illuminate psychobiological mechanisms associated with social and personality processes in future research.

**The Future of Social Neuroscience in Social-Personality Psychology**

Over the past decade, physiological approaches have re-emerged as an important facet of social-personality research, now augmented by advances in neural, pharmacological, endocrinological, immunological, and genetic approaches. This time, we think neuroscience is here to stay. Recently a novelty in social-personality circles, the neuroscience perspective is now woven into the natural discourse of social-personality inquiry. Neuroscience data are increasingly integrated into the literature reviews of mainstream social-personality manuscripts, and psychophysiological methods complement the traditional tools of behavioral psychology, now without the fanfare of a novelty act. Social neuroscience is also becoming more prevalent in the
training of new social-personality psychologists, not to mention other disciplines of psychology or allied fields such as economics and communication. These are healthy developments for the field.

The purpose of this volume is to highlight the interplay of personality and social psychological approaches. It is interesting to consider this aim from the perspective of social neuroscience, a field in which traditional boundaries between the person and the situation are reinterpreted as complex, dynamic, and inherently multi-level interactions. For example, neuroscientific models reveal how our perception of a situation is influenced by dispositional factors, such as personality, goals, and mental sets (following ideas from the New Look movement). At the same time, research on genetics shows that even our DNA may be influenced by situational factors at the time of conception, and that gene expression – often held to be the purest expression of personality – is strongly influenced by the situation. Thus, from the social neuroscience perspective, a dynamic interplay of personal and situational influences operates at every level of analysis.

In this regard, the social neuroscience approach is helping to build connections between the fields of social and personality psychology in two ways. The first way is through its influence on theory and research. As noted above, social neuroscience research reveals the dynamic symbiosis between situational and personal factors that exists across levels of analysis, in line with the interactionist view that the effects of personality and the situation can only genuinely be studied in the context of each other. The second way is by bringing together researchers from different disciplines to lend their respective expertise to integrative research questions, and promoting education and training in interdisciplinary approaches used in social neuroscience research.
References


Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A.R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and


Hare, T., O'Doherty, J., Camerer, C., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of Neuroscience, 28*, 5623-5630.


insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology, 80*, 797-803.


Handbook of Personality Theory and Research (pp. 251–276). New York: Guilford.


Salamone, J. D. (2007). Functions of mesolimbic dopamine: changing concepts and shifting...


and Social Psychology, 62, 676-687.


Vyta, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic


Acknowledgments

Work on this article was supported by Grants from the National Science Foundation to David Amodio (BCS 0847350) and Eddie Harmon-Jones (BCS 0643348, BCS 0921565), and from the National Institutes of Health to Elliot Berkman (CA175241 and AG048840).
Figure Captions

1. Diagram of amygdala illustrating relations between the lateral nucleus (LA), central nucleus (CeA), basal basal nucleus (BA), and intercalated masses (ITC) and their respective functions. Adapted from Amodio (2014).

2. Medial aspect of the left hemisphere of the brain. mPFC = medial prefrontal cortex, OFC = orbital frontal cortex.

3. View of coronal slice through brain, with structures on the left side labeled. AMG = amygdala.

4. Lateral aspect of the right hemisphere of the brain. dLPFC = dorsolateral prefrontal cortex, vLPFC = ventrolateral prefrontal cortex, pPFC = posterior prefrontal cortex, TPJ = temporo-parietal junction.
Figure 4

![Diagram of brain with labeled regions: TPJ, dIPFC, pPFC, vIPFC, Lateral temporal lobe, and Fusiform.](image)