

Impression Formation:
A Focus on Others' Intents

Daniel L. Ames

Susan T. Fiske

Alexander T. Todorov

Princeton University

Keywords: impression formation, social cognition, attribution, theory of mind, intent, capability, dispositions, agents, warmth, competence, trustworthiness, dominance, face perception, mPFC, amygdala, Stereotype Content Model, fMRI, social neuroscience, social perception, neuroimaging

Chapter to appear in J. Decety & J. Cacioppo (Eds.), *The Handbook of Social Neuroscience*.

Oxford University Press

Living things need to predict their environments in order to survive. This fundamental fact of nature appears in the impressive array of predictive adaptations observed across species. For example, the archer fish, which feeds by shooting insects off branches with jets of water, predicts the 3D trajectory of its dislodged prey, allowing it to move to the splash point before its meal (or another fish) arrives (Rossel, Corlija & Schuster, 2002). Similarly, some species of flies predict (with frustrating accuracy) the most likely trajectory of a looming flyswatter and plot an optimal escape path before the threat descends (Card & Dickinson, 2008).

While the computational challenges posed by these everyday feats of prediction are incredibly complex, the daily predictive challenges faced by human beings are orders of magnitude more complicated. This is largely because what people most need to predict in order to obtain their desired outcomes is other people—agents whose actions arise from internal, volitional causes. Because they are intentional agents, people are notoriously difficult to predict, and often prefer it that way. How then do we predict the willfully unpredictable?

Attribution theories (e.g., Heider, 1958; Jones & Davis, 1965; Kelley, 1967) suggest that people understand each other's behavior as arising from two causes: dispositions (i.e., personalities) and situations. From this perspective, predicting how intentional agents will behave can be accomplished via simple algebra: disposition + situation = behavior (Lieberman, Gaunt, Gilbert, & Trope, 2002). By combining dispositional information about Kevin (who has a morbid fear of pachyderms) and situational information (an elephant is traipsing into the room), one should be able to make reasonable predictions about what behaviors Kevin might enact. In this example, the relationship between situation and disposition is relatively straightforward; however in the real world, the kaleidoscopic interplay between these two factors is fantastically complex (perhaps irreducibly so; Gilbert, 1998), making it frustratingly difficult to implement the clean logic of the attributionist's equation.

Luckily, people have developed a remarkably efficient way of simplifying the calculus: They ignore the elephant in the room. A voluminous body of research has demonstrated that people tend to neglect situational information when trying to make sense of others' actions (e.g., Gilbert & Malone, 1995; Jones & Harris, 1967; Ross, 1977). One consequence of this is that when people undertake the difficult but crucial job of predicting what other agents will do, the dispositional impression is the tool of choice. That it is not always the best tool for the job does not limit people's use of it. Like other blunt instruments, dispositional inferences are easy and satisfying to apply; and they promise results (albeit imprecise ones), only occasionally causing irreparable damage in the process. In their more assiduous moments, people will complement their stable impressions of others with situational information, combining the dispositional hammer with the fine edge of a situationist's chisel to carve out detailed predictions and interpretations of others' behavior (Gilbert, Pelham, & Krull, 1988; Trope & Gaunt, 2003). Usually, however, people are contented to swing hammers freely and leave the chisel to rust. Because dispositional impressions are the tool of choice for predicting other people, impression formation plays an enormous role in determining how people think about and navigate the social world. As such, understanding the cognitive processes involved in forming these impressions has long been a primary objective of social cognition research.

The advent of social neuroscience provides a new avenue for understanding social impression formation. Capitalizing on recent technological advances, researchers have begun to characterize how the brain gives rise to coherent, stable impressions of other people, despite the complexity of people's behavior. One of the major advantages of this approach is that it affords unique opportunities for revealing areas of convergence and dissociation across cognitive processes.

Overview

This chapter reviews research on social impression formation, focusing specifically on how social neuroscience has contributed to our understanding of this complex but fundamental social process. For convenience, this review organizes around three different ways that people form impressions of one another: secondhand information (being told about someone), direct behavioral experience (interacting with someone), and appearance (seeing someone's looks). While the lines between these three information sources often blur, studies focusing on one kind or another have tended to elicit different patterns of neural activity, with the more deliberative tasks involving secondhand information and direct experience most frequently recruiting medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), and posterior regions of superior temporal sulcus (pSTS), and tasks involving more automatic, appearance-based judgments most frequently recruiting the amygdala. The chapter concludes by comparing impression formation and intentional inference, and then discussing their functional relationship.

Impression Formation via Secondhand Information

Human beings are motivated to learn about others. Indeed, our appetite for social knowledge seems to far exceed the limits of what we can absorb through direct experience. Some 65% of people's conversational time is devoted to social topics (Dunbar, Marriott & Duncan, 2007). Consequently, social impressions often result from secondhand information. This section reviews research on the neural processes involved in forming impressions from what perceivers are told about other people (as opposed to impressions based on others' appearance or actions—topics covered in later sections). Presenting participants with this kind of secondhand information about other people has been the most popular way of investigating impression formation processes in social neuroscience—in part because it is easy to implement. Because a wide range of studies have employed this approach, this section subdivides into four different topics. The first subsection explores how the marriage of social neuroscience with

classic attribution theories has generated insights about the brain regions involved in impression formation. The second subsection considers whether social impression formation relies on different neural substrates than other kinds of impression formation. The third subsection asks how impression formation changes subsequent neural responses to other persons. The final subsection considers whether these changes reflect content-specific representations or only general evaluations.

Attribution In The Brain

Attribution theories of person perception (Heider, 1958; Jones & Davis, 1965; Kelley, 1967) represent one of the longest and richest theoretical traditions in social cognition research, and this tradition offers many well-validated paradigms for studying how people form impressions of one another. A recent study adapted one such paradigm (McArthur, 1972) to examine what specific kinds of information trigger the neural mechanisms that give rise to dispositional attributions (Harris, Todorov, & Fiske, 2005). McArthur's original study (1972) sought to test the prediction that specific patterns of information about people's behaviors (consensus across actors, distinctiveness across entities, and consistency over time) induce people to make dispositional inferences (Kelley, 1967). For example, if Mike laughs at the comedian, observers may not know why; but if they also know that no one else laughs (low consensus), that Mike laughs at every comedian (low distinctiveness), and that he always laughs at this comedian (high consistency), then observers will readily deduce that Mike is easily amused (a dispositional inference).

McArthur's (1972) participants read descriptions of various actions, along with eight combinations of information about those actions related to high and low consensus, distinctiveness, and consistency. Participants then indicated whether they thought each action (e.g., Mike laughing) was most likely caused by (a) something about the person (Mike), (b)

something about the stimulus (the comedian), (c) something about the particular circumstances (Mike was drunk at the time), or (d) some combination thereof. As Kelley's model predicts, when actions appeared as low-consensus, low-distinctiveness and high-consistency (as in the example involving Mike and the comedian), participants overwhelmingly chose option (a), that is, they made a dispositional inference, an impression that explained his actions.

This behavioral result was replicated when participants completed a computerized version of the task during fMRI scanning (Harris, Todorov, & Fiske, 2005). Moreover, only this specifically dispositional combination (and none of the other seven McArthur combinations) activated posterior STS (previously implicated in perceived intent and intentional trajectories; e.g., Gobbini, Koralek, Bryan, Montgomery & Haxby, 2007; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). Decades of behavioral research show, however, that people often ignore information about consensus across actors (whether only Mike laughs) (see Fiske, 2004 for review). Ignoring consensus information in this study, both the high-consistency/low-distinctiveness combinations activated mPFC above baseline, whereas the remaining six combinations did not. This work converges with research reviewed in subsequent sections, which shows that STS and mPFC play crucial roles in impression formation across a variety of tasks.

Different Systems for Secondhand Person and Object Knowledge

The introduction offered two related suggestions about the nature of social impression formation. First, because social agents' actions prominently arise from dispositions, which are *internal* causes, while nonsocial actions arise from *external* causes, understanding and predicting other people involves fundamentally distinct challenges from understanding and predicting things (e.g., bugs, flyswatter trajectories). Second, people use social impression formation as a way to meet the particular challenges involved in predicting intentional agents. These

suggestions together imply that, although the term “impression formation” could describe understanding many sorts of things—couches and musical genres, as well as people—social and nonsocial impression formation are, in fact, distinct mental events with dissociable underlying processes.

fMRI investigations of social vs. nonsocial impression formation support this distinction. For example, in one study (Mitchell, Macrae, & Banaji, 2005), participants read about both people and objects during fMRI scanning and either formed an impression of each target or performed a (perceptually identical) “sequencing” task, memorizing the order in which the information about each target appeared. Dorsal aspects of medial prefrontal cortex (dmPFC) were more active for the social impression formation task than for all other conditions (including forming impressions of objects and memorizing the order of information about people). The mPFC has been heavily implicated in a wide array of social cognitive tasks (for meta-analyses, see Amodio & Frith, 2006; Gallagher & Frith, 2003; Van Overwalle, 2009). Thus, the preferential activation of mPFC for forming impressions about people suggests a particular role for this region in the social-cognitive components of impression formation (we hasten to add that mPFC is a large area of cortex, with various subregions involved in many kinds of processing besides social cognition; e.g., Duncan & Owen, 2000; Schacter, Addis & Buckner, 2008. While a complete review of the mPFC lies beyond this chapter’s scope, some alternative theories of mPFC function are discussed toward the end of this chapter).

Even stronger support for a social/nonsocial processing distinction appears in a similar study (Mitchell, Macrae, & Banaji, 2004). In this experiment (which used the same social impression formation and relatively nonsocial “sequencing” conditions), dmPFC again engaged more for impression formation than for sequencing. A strikingly different set of regions was more active for the sequencing task than for the impression-formation task (including the caudate

as well as superior frontal, parietal and precentral gyri). This double-dissociation was echoed by a second double-dissociation in participants' memory for information in both conditions: The better participants' memory for information presented in the impression-formation condition, the more dmPFC activated during the impression formation task (but not during the nonsocial sequencing task). In contrast, memory performance for the sequencing task (but not impression formation) correlated with right hippocampus activity (a region more generally involved in memory; Squire, 1992). These memory findings imply that mPFC may be preferentially involved in a distinctly social form of processing, a suggestion corroborated by demonstrations that dmPFC activity is specifically predictive for the encoding of social versus nonsocial pictures (Harvey, Fossati, & Lepage, 2007).

Thus, while social impression formation may specifically link to dmPFC (Mitchell et al., 2005; Van Overwalle, 2009), the results could also stem from a more general social/nonsocial processing distinction. Consistent with this idea, other studies have demonstrated a division of neural architecture for applying knowledge about agents vs. objects (with mPFC subserving agent knowledge; Mason, Banfield, & Macrae, 2004; Mitchell, Heatherton, & Macrae, 2002). Because such bifurcated systems are both anatomically and metabolically expensive, these findings raise the question of why these divisions might exist. One possibility is that such a system confers advantages in speed and accuracy of information processing. When different categories of knowledge are subserved by common neural architecture, competition between categories may cause interference, thus compromising information processing (Mason et al., 2004, see also Caramazza, 2000; Otten & Rug, 2001). Such interference can have dire consequences in evolutionarily important domains that require the rapid translation of knowledge into action. For human beings, social cognition is surely one such domain.

Secondhand Impression Formation Changes Subsequent Person Representations

Indeed, for impression formation to be useful, it has to change how people act toward others. If rumor has it that one cinema ticket-booth operator is perpetually surly, rarely showers, and refuses to serve every fourth customer, while the other is generally cheery, practices good hygiene, and gives away high-end electronics with each purchase, moviegoers will want to make sure to queue in the right line. To distinguish the two employees, one must link the crucial information to the appropriate person, making it accessible for later use (e.g., Goren & Todorov, 2009; Todorov & Uleman, 2002, 2003, 2004). The next experiments examine the neural substrates of this process.

In one revealing study (Delgado, Frank, & Phelps, 2005), participants read detailed information about the life events of three individuals, suggesting exemplary, neutral, or dubious moral character. Later, they played a trust game with each of these partners while undergoing fMRI scanning. Behavioral and neural data suggested that, when diagnostic information was available (in the “good” and “bad” partner conditions), participants tended to rely less on their opponents’ actual behaviors in the game to predict how they intended to play future rounds—instead using their prior impressions to predict others’ intentions. Behaviorally, participants persisted in trusting the “good” partner to play cooperatively, despite equivalent payouts for all trading partners. The fMRI findings showed that activity in the caudate nucleus (an important structure in reward learning; Poldrack et al., 2001) distinguished whether the outcome of each round was positive or negative for the participant, consistent with past work on neural reward-feedback systems (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000). However, this pattern appeared robustly only for the neutral partner, about whom no diagnostic information for impression-formation had been presented. Prior impressions of the “good” and “bad” partners apparently reduced reliance on the neural reward-feedback systems involved in trial-and-error learning. This result converges with the suggestion that impression formation is people’s

preferred method of predicting others' intentions by showing that, when people have the opportunity to use dispositional information about another person to predict intentions, they seem to use that information rather than actual behavior to guide their forecasting of what others will do. Although scores of behavioral studies have indicated people's over-reliance on social expectancies (Fiske & Taylor, 2008), these fMRI data provide clues regarding the distinct brain systems potentially involved.

While maintaining stable impressions as predictive models of other people in the face of inconsistent behavior may sometimes lead to suboptimal outcomes (in this case, over-investing in the "good" partner in an economic game), maintaining a robust impression is an adaptive strategy overall. Updating our long-term predictive models of other people following each instance of behavior could interfere with ongoing activity (McClelland, McNaughton & O'Reilly, 1995), preventing the use of impression formation as a long-term predictive strategy. Because people are agents, they do not always behave as expected, even when expectations are sensible. For example, even if we know that chocolate is Karen's favorite flavor of ice cream, she will probably still choose other flavors with some frequency. While being too quick to write off our diagnosis of Karen as an inveterate chocoholic may not have dire consequences, the matter becomes much more serious when one considers other situations, such as how heavily to weigh evidence of recent good acts in gauging whether an individual with a history of violent crime is likely to aggress again.

How much prior information is required in order to establish an impression that alters subsequent neural responses to others? In the trust game experiment (Delgado et al. 2005), participants had access to fairly detailed information about each target prior to scanning. However, a more recent study (Todorov, Gobinni, Evans, & Haxby, 2007) based on prior behavioral experiments (Todorov & Uleman, 2002, 2003) demonstrates that, at least in some

instances, a single piece of information will suffice. In the first phase of this study, participants viewed faces paired with a statement about that person's (ostensible) previous behavior. In the second phase, participants underwent fMRI scanning while viewing the faces from phase one intermixed with a set of other faces. The task that participants performed in the scanner (deciding whether or not each face was the same as the one preceding it) was perceptual and did not require the retrieval of social knowledge. Nevertheless, faces previously paired with behaviors evoked stronger responses in STS (as noted, often associated with inferred intent) and dmPFC than did novel faces, suggesting the activation of impressions irrelevant for the current task. Perhaps more surprisingly, responses to specific faces were influenced by the type of behavior with which each face had previously been associated. For example, compared with faces previously paired with aggressive behaviors, faces previously paired with disgusting behaviors reliably elicited greater response in anterior insula, a region involved in processing disgust (Philips et al., 1997). These studies suggest that impression formation does meet the usefulness criterion laid out at the beginning of this section: Because impressions change how people predict others' intentions (Delgado et al., 2005), and even simply perceive those others (Todorov et al., 2007), social impressions can usefully guide social behavior.

Content-Specific Representations or Mere Valence?

Merely thinking about or seeing another person suffices to make accessible previously-learned information about that person. But whether such activations encode content-specific impressions or only general evaluations has, until recently, remained unclear. This ambiguity sparked an investigation of the neural systems responsible for the general evaluative component of first impressions (Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009). During scanning, participants viewed a series of target individuals and six pieces of information about each person that varied in affective valence. In separate trials, participants also viewed each face without

information. Participants then gave their overall evaluative impression of each target on a 1-8 scale. The dmPFC and several other areas were more active when the faces accompanied information about the target than when the faces of the targets appeared alone, providing additional evidence for the specific recruitment of dmPFC for social impression formation, as opposed to the mere perception of persons (Amodio & Frith, 2006; Mitchell et al., 2004). However, dmPFC did not relate to participants' evaluations of the targets, while amygdala and posterior cingulate cortex (PCC) regions (frequently implicated in emotion and valuation processes) did relate (Cunningham et al., 2008; Kable & Glimcher, 2007; Taber et al., 2004). Specifically, parametric analyses revealed greater activity in amygdala and PCC during the viewing of positive information when participants evaluated targets positively, and during the viewing of negative information when participants evaluated targets negatively. That is, regardless of any given participant's idiosyncratic impressions, these two regions reliably tracked the affective components of their impressions. No such pattern emerged for dmPFC. Consistent with prior theories (Firth, 2007), the role of mPFC in impression formation appears dissociable from the general affective functions of evaluative subsystems in the brain—with the mPFC perhaps playing a special role in encoding higher-order representations of others' mental lives.

A Cautionary Note

Despite the growing evidence for the involvement of mPFC and STS in impression formation culled from studies employing written descriptions of people, this format may exaggerate people's dispositional inferences and therefore provide a biased depiction of the brain systems that covary with these inferences. For example, behavioral experiments show that secondhand information under-emphasizes situational qualifications and results in more severe judgments (Gilovich, 1987), and leads to stereotypes more extreme than those learned via direct contact (Thompson, Judd, & Park, 2000). Hence, the next section turns to neural processes

involved in impressions based on direct experience.

Impression Formation via Interaction

Reading about others is a useful way to learn about people. However, notwithstanding the best efforts of Internet search companies and social networking websites, people continue to gain some proportion of their knowledge about others by interacting with them. While interaction is perhaps the most obvious way to form impressions of other people, the neural processes underlying interaction-based impression formation have been relatively under-studied to date. Emphasis on secondhand information characterizes social cognition research more generally (Fiske & Taylor, 2008); however the current limitations of functional neuroimaging research (e.g., the requirement that participants in fMRI studies lie motionless and alone inside a dark, noisy bore) make studying actual social interactions at the neural level particularly challenging.

Investigators have negotiated these impediments by simplifying social interactions to fit the scanning environment. Often, the interactions take the form of economic games. Though artificial by design (indeed, their constrained nature is what makes them useful), economic games permit participants to engage in socially meaningful actions, such as trusting, cooperating, and betraying.

Success in these games depends largely on participants' ability to predict their opponents' intentions (e.g., to cooperate or defect on the next round; Delgado et al., 2005; Van Overwalle, 2009). Consistent with the idea of impression formation as a tool for predicting others' intentional behaviors, players appear to meet the predictive challenges of economic games by forming impressions of their opponents based on their actions in previous rounds of the game (unless they have already formed an impression by other means; Delgado et al., 2005; reviewed above). These impressions reflect neural responses to opponents' faces after their interaction. In

one study (Singer, Keibel, Winston, Dolan, & Frith, 2004), participants underwent fMRI scanning while judging the gender of opponents who had previously played fairly or unfairly in an ultimatum game, and likewise judged the gender of other targets for whom participants had no behavioral basis for impression formation (matched for earlier exposure). Despite the irrelevance of impressions to gender judgments, faces of participants' previous opponents in the ultimatum game elicited greater activity in amygdala, orbitofrontal cortex, and anterior insula, regions that have frequently been implicated in forming rapid, intuitive, judgments of others' trustworthiness and approachability (Damasio, 1994; Adolphs, 2003; Winston, O'Doherty & Nolan, 2003; Todorov, in press).

Note that these activations occurred subsequent to participant-opponent interactions, outside the context that informed the impressions. Thus, impressions as models for predicting others' intentions show at least some contextual invariance—that is, people appear to use impressions to predict one another even when those impressions are formed under specific contexts that may not approximate how people will behave outside of those contexts. As suggested earlier, people rely on impression formation to predict others because people are agents: They choose to do things. Hence, contextual insensitivity in impression formation may arise partially from participants' sense (either tacit or explicit) that the opponents themselves, rather than contextual constraints, controlled their behavior. To examine this possibility, participants in the Singer et al. (2004) study were told that some of their opponents were simply following a computer's instructions rather than freely choosing their responses. As expected, the intentionality of opponents' actions impacted subsequent neural responses to those opponents' faces; for example, intentional cooperators elicited greater recruitment of OFC and posterior STS than did non-intentional cooperators. As Frith and Frith (2006a) observed, “[s]ubjects were not simply learning which faces were associated with reward. They were learning whom to trust” (p.

38).

A second experimental game study demonstrates the relationship between experience-based impression formation and predicting others' intentions even more directly (King-Casas et al., 2005): Behavioral and neural responses to others changed as people became acquainted during an economic decision-making game. As in other studies, neural signals in the caudate nucleus apparently encoded reward value. Within the context of a trust game, where rewards related to trusting other players, this caudate reward signal reliably predicted the intention to trust one's opponent. As participants gained experience with their opponents, this "intention to trust signal" (King-Casas et al., 2005) became anticipatory, shifting earlier by 14 seconds over the course of the study. This suggests that participants were predicting their partners' intentions in the next round further in advance as they got to know their opponents. In support of this interpretation, participants became increasingly accurate in predicting other players' intentions over the same period during which this shift in neural activity took place. Cross-brain analyses examining correlations between players in "trustee" and "investor" roles yielded a strong correlation between ventral aspects of mPFC in the trustee's brain and the middle cingulate of the investor's brain (which most strongly activated when the investor made a decision). The development of such cross-brain analysis techniques (see also Hasson, Nir, Levy, Fuhrmann & Malach, 2004) may facilitate applying neuroscience to dynamic social processes. Altogether, this study demonstrates how forming impressions of others through interaction results in neural changes that allow people to predict more accurately others' future intentions.

A recent survey of several other economic game studies (Van Overwalle, 2009) drew similar conclusions about the role of dispositional inferences in predicting partners' intentions. Fully 100% of economic game studies reviewed in this meta-analysis yielded significant activations in regions of mPFC that have been strongly implicated in social cognition, with 86%

of these activations occurring in dmPFC, the region of the brain most consistently implicated in social impression formation and ascribing dispositional traits to others (e.g., Mitchell et al., 2004; 2005; Van Overwalle, 2009). The same meta-analysis observed frequent activation of STS in conjunction with the presentation of biological motion, such as human movement and glancing behavior (with such motion often being related to intentional inference; see also Macrae & Quadflieg, in press).

Although experimental games cannot (and are not intended to) recreate naturalistic human interactions, they provide social neuroscience a way to study how interpersonal experience influences social impressions. In addition, because these games motivate players to predict their opponents' intentions, they offer a particularly appropriate context for studying the relationship between impression formation and interpersonal prediction. Together, the studies reviewed in this section illustrate the neural processes through which social interactions can shape interpersonal impressions, and how these impressions, in turn, help people predict others' intentions.

Impression Formation via Appearance

In a counterpoint to the first- and secondhand impressions research, people adeptly form impressions without benefit of any behavioral information at all. The human face offers a powerful social stimulus; a fleeting glance at another's face provides a wealth of information (and misinformation) about others' transient states and stable dispositions (Macrae & Quadflieg, in press). One key finding: People judge another person's trustworthiness based on facial appearance after as little as 33 ms of exposure (Todorov, Pakrashi, & Oosterhof, in press). Although such judgments are not necessarily accurate, they nonetheless predict important social outcomes, including criminal sentencing and success in being elected to public office (Ballew & Todorov, 2007; Todorov, Mandisodza, Goren & Hall, 2005, see also Eberhardt, Davies, Purdie-

Vaughns & Johnson, 2006).

Impressions based on appearance alone tend to be rapid, intuitive, and emotional. As such, this research contrasts with the previously-reviewed studies, which lean toward slower, more deliberative impression formation processes that build on more explicit social information. Whereas mPFC and STS tend to be the most frequently observed regions in these deliberative, explicit processes, the amygdala appears to matter more for intuitive judgments based on appearance.

Perhaps the most-studied appearance-based impression in social neuroscience has been trustworthiness (Todorov, in press). Trustworthiness appears as a universal dimension of person perception, correlating with many other important social judgments (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008; see also Fiske, Cuddy, & Glick, 2007, below). As such, trustworthiness assessment is crucial to impression formation. Neuroimaging studies implicating the amygdala in trustworthiness judgments fit neuropsychological evidence of patients with bilateral amygdala damage, who exhibit impaired performance in discriminating between trustworthy- and untrustworthy-looking faces (Adolphs, Tranel & Damasio, 1998). Several researchers have reported increased amygdala responsivity to faces judged untrustworthy (e.g., Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002), as well as to members of racial outgroups (Hart et al., 2000; Phelps et al., 2000). Such findings align with initial views of the amygdala as a fear/threat-detection region (Phelps, 2006).

However, recent evidence suggests that the amygdala exhibits a nonlinear response pattern, activating preferentially to both highly trustworthy and highly untrustworthy faces (Said, Baron, & Todorov, 2009). This finding fits other observations of amygdala activation correlating with attitudinal intensity, rather than valence per se (Cunningham, Raye, & Johnson, 2004; Cunningham et al., 2008). One intriguing possibility is that the amygdala encodes, not threat or

fear, but vigilance, directing attention to emotionally important information in one's environment (Vuilleumier, 2005). Thus, within the context of social impression formation, increased amygdala activation may help direct attention toward people who appear particularly likely to help or harm—perhaps by providing arousal cues coinciding with the perception of those individuals. Increased arousal and attention toward likely friends and foes would facilitate the efficient encoding of their behaviors. In other words, paying closer attention to what people do provides a stronger informational basis for detailed impression formation and, by extension, for the accurate prediction of behavior. The amygdala (working in conjunction with other regions) may help to serve this function by biasing attention toward particularly trustworthy and untrustworthy individuals (Adolphs, 1999; Haxby, Hoffman, & Gobbini, 2000).

This suggestion fits Todorov's recent theoretical proposal that judgments of facial trustworthiness derive from overgeneralizing facial cues signaling emotional valence (i.e., anger and happiness; Todorov & Engell, 2008). Knowing whether another person is angry or happy is useful for a number of reasons, not least because it indicates whether another's intentions are likely to be friendly or hostile. From this perspective, human beings would be highly attuned to facial features that signal these states—perhaps even to the extent that permanent features of facial geometry (e.g., low eyebrows) could imply a subtle emotional message (e.g., anger), causing the perceiver to make a dispositional inference about the target (e.g. untrustworthiness). Several studies have confirmed that even subtle manipulations of facial cues encoding happiness and anger are sufficient to impact judgments of dispositional trustworthiness (Oosterhof & Todorov, 2008). The emotion overgeneralization hypothesis (Montepare & Dobish, 2003; Oosterhof & Todorov, 2009; Said, Sebe, & Todorov, 2009) gains further support from the observation that nonlinear patterns of amygdala responses to facial trustworthiness (Said, et al., 2009; Todorov, Baron, & Oosterhof, 2008) closely mirror earlier observations that the amygdala

is more responsive to both happy and fearful faces than to neutral faces (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002).

Thus, people are highly sensitive to social information conveyed by facial appearance, and facial information serves as a powerful basis for rapid and intuitive impression formation. Most likely, such impressions are driven largely by extrapolation from physiognomic features encoding subtle emotional messages that are interpreted (perhaps misinterpreted) as cues about a person's trustworthiness. A second dimension of face-perception, dominance, has been identified as orthogonal to trustworthiness judgments. Just as trustworthiness judgments appear to be extrapolations from valence cues, perceptions of facial dominance appear to be driven by subtle facial features encoding masculinity and femininity (Oosterhof & Todorov, 2008). Presently, the neural correlates of dominance perception are less well understood than those pertaining to judgments of facial trustworthiness; however important social judgments (e.g., how threatening a person is) can be modeled as conjunctions of trustworthiness and dominance (Oosterhof & Todorov, 2008) (see Figure 2).

Besides faces, people use other aspects of appearance to judge others. Behavioral studies demonstrate the universality of two primary dimensions, accounting for about 80% of the variance in first impressions (Fiske, Cuddy, & Glick, 2007). Consistent with the Todorov et al. studies, the most rapid and apparently primary dimension assesses people's intentions as warm (trustworthy, friendly, sincere, moral, communal) or not. This dimension is predicted by cooperative or competitive relationships, such as those implicated in the experimental game literature. The second dimension reflects people's ability to enact their intentions, that is, competence (capability, skill, agency). This dimension is predicted by status, consistent with the Todorov results for perceived facial dominance. This robust behavioral literature uses stimuli ranging from personal acquaintances, to videos and photographs, to stereotypes and expectations

(see Figure 3).

Recent neuroimaging studies investigating the dimensions of warmth and competence have capitalized on people's willingness to form impressions based on appearances. In these studies, participants view photographs of people who are easily identified, based on their appearance, as representing the four distinct combinations of people high and low on warmth and competence. One series of studies especially implicates the by-now familiar mPFC in perceptions of all combinations except one: those perceived as both low-warmth and low-competence. The lowest of the low, essentially outcasts, are poor people, especially those who appear to be homeless, as well as people apparently drug-addicted. Photographs of these individuals uniquely fail to activate mPFC above baseline (Harris & Fiske, 2006), and behavioral data fit the interpretation that observers tend to dehumanize the outcasts, reporting difficulty with perceiving their minds: intentions, thoughts, and feelings (Harris & Fiske, in press). As converging evidence for the role of the mPFC in forming social impressions, an instruction to consider the preferences of the pictured individuals suffices to bring the mPFC back on-line (Harris & Fiske, 2007).

In sum, appearances can inform impressions. In contrast to studies examined in earlier sections, which focused on relatively deliberative inferences based on relatively substantive information, the work reviewed in this section shows that people can form rapid and intuitive impressions based on almost no information at all. People readily extrapolate personality features from facial features, overgeneralizing features that encode socially relevant emotions. People also make direct assessments of primary social dimensions based on appearance; and these assessments partially determine whether the neural mechanisms used for thinking about other minds come online. Thus, impressions formed on the basis of minimal social information can have consequences for social interaction.

Impression Formation and Intentional Inference: Theoretical and Functional Overlap

This chapter has examined three kinds of information that people use to form impressions: secondhand information, direct experience, and appearance. Across these three domains, various studies have supported the theoretical relationship suggested at the chapter's outset between impression formation and predicting others' intentions. This final section explores this connection at a functional level, capitalizing on the ability of social neuroscience to identify areas of overlap and dissociation in cognitive processes.

The suggestion that there may be a functional relationship between impression formation and intentional prediction may seem misguided, because the neural literature often sharply dissociates these two processes, with intentional inference ascribed principally to posterior regions of STS and impression formation usually associated with mPFC. Like most clear-cut neural distinctions and pre-mixed peanut butter and jelly, this proposal sacrifices something worthwhile for the sake of convenience. At least three sets of findings support a more nuanced story in which the mPFC contributes to the representation of others' intentions (though the STS certainly does so as well).

First, intentional inference is not a monolithic mental operation and, therefore, not likely supported by a single set of neural substrates. Social neuroscience usefully distinguishes the representation of intentions at a *perceptual* level from the representation of *covert mental states* that may predict future intentional actions; posterior STS contributes to perception of intention, while mPFC and temporo-parietal junction (TPJ) serve the inference of intention (Gobbini, et al., 2007; Saxe & Powell, 2006; Saxe et al., 2004). That is, inferring intentions from present, perceptible actions (e.g., biological motion; Saxe et al., 2004) is, to some extent, a qualitatively different process compared with inferring intentions from tacit information that is not presently perceptible (e.g., another's beliefs, Young & Saxe, 2008).

Second, mPFC shows an inconsistent pattern of engagement across intention-relevant

tasks, perhaps due to a selective functional profile for different *kinds* of intention. Specifically, present evidence suggests that mPFC may be critical for representing others' *social* intentions (for example, the intention to communicate with another person), but not other kinds of intentions (Ciaramirado, et al., 2007; Kampe, et al., 2003; Walter, et al., 2004). In contrast, STS may represent both social and nonsocial intentions.

Third, the relative contributions of STS and mPFC to thinking about intentions vary as a function of participant age. Although both adolescents and adults recruit both regions in intentional inference tasks, adolescents rely more heavily on mPFC, while adults show greater recruitment of STS (Blakemore, et al., 2007). This suggests that, while most normally developing children are able to pass theory-of-mind tasks (which typically test the ability to reason about others' beliefs) by about age 5 (Barresi & Moore, 1996), the computational strategies used to think about intention continue to be refined at least into early adulthood. In sum, both mPFC and STS contribute to the understanding of intention, with their relative involvement varying as a function of the level at which the intention is being inferred (perceptual vs. covert), the type of intention (social vs. nonsocial) and the age of the perceiver.

Of course, mPFC is a large area of cortex, and has been implicated in many different social functions (Amodio & Frith, 2006; Frith & Frith, 2001, 2006a, Van Overwalle, 2009). Note, however, that at least some studies suggest that there may be considerable neural overlap in the regions involved in impression formation and intentional prediction. For example, the region reported by Mitchell et al. (2004; described above) for impression formation bears a striking resemblance to the activation reported by Walter et al. (2004) for the prediction of social intentions. Both sets of findings have been replicated (impression formation: Mitchell et al., 2005; predicting others' intentions: Ciaramirado et al., 2007). Although caution must accompany comparing results across studies, this overlap is echoed in meta-analyses (Amodio &

Frith, 2006; Van Overwalle, 2009). The question of whether and to what extent these two processes truly rely on the same neural architecture awaits an empirical test that examines both intentional inference and impression formation in the same study.

Other studies have identified a similar region as playing an important role in imagining the future (Addis, Wong & Schacter, 2007; Schacter, Addis & Buckner, 2007; 2008). Perhaps, within the context of social cognition—which is subserved by a network of regions including mPFC, TPJ, STS, medial parietal cortex, and temporal poles (Fletcher et al., 1995; Gallagher et al., 2000; Gallagher, Jack, Roepstorff & Frith, 2002; Goel, Grafman, Sadato, & Hallett, 1995; Saxe & Kanwisher, 2003; Van Overwalle, 2009)—mPFC may serve a special function in predicting others' future mental states, the better to anticipate their actions (Frith & Frith, 2006b). Such observations appear consistent with the findings reviewed herein demonstrating the important role played by impression formation in predicting others' intentions. Note, however, that this region is also implicated in remembering the past (Addis, et al., 2007). Thus, the involvement of mPFC in inter-temporal construction could be entirely separable from its involvement in impression formation. Alternatively, the two processes might converge in cases where impressions formed based on past experience are marshaled for predicting one's own or others' future intentions and behaviors (see also Buckner & Carroll, 2006).

Note that this chapter does not claim the mPFC is selective for impression formation processes or even for social cognition. Indeed, as noted, mPFC is a large, complex region of cortex with a host of different functions (Amodio & Frith, 2006). For instance, subregions of mPFC (particularly more ventral/orbitofrontal aspects) have been consistently implicated in reward processing (Montague, King-Casas & Cohen, 2006) (though even some of these subregions are differentially responsive to social vs. nonsocial rewards; Harris, McClure, van den Bos, Cohen & Fiske, 2007; van den Bos, McClure, Harris, Fiske & Cohen, 2007).

Nonetheless, the evidence for the involvement of mPFC in social cognitive processes including impression formation and intentional prediction is, by now quite substantial (e.g., Amodio & Frith, 2006; Gallagher & Frith, 2003; Van Overwalle, 2009). While the processes subserved by mPFC may not all be specifically social, because they can demonstrably be recruited for use in other domains, it may be that the evolutionary importance of complex social tasks has been largely responsible for causing these processes to develop in increasingly sophisticated ways (Frith, 2007).

The proposed relationship between impression formation and intentional inference outlined in this chapter bridges between what have sometimes been seen as two separate components of social cognition: inferences of transitory states and inferences of enduring characteristics (Frith & Frith, 2006a; Van Overwalle, 2009). The research reviewed here points to functional overlap between these two processes, highlighting a strength of neuroscientific approaches to social cognition research: the ability to determine whether and to what extent two apparently disparate social cognitive responses involve fundamentally similar neural processes. This capacity to force theoretical convergence meets an equally useful and opposing property: the ability to determine when two apparently similar behaviors are driven by different cognitive processes. Overall, the ability of social neuroscience to measure the extent of overlap and dissociation between processes may shed light on both novel questions and longstanding theoretical debates about social impression formation. Just as impression formation is a critical tool in understanding other people, so too is social neuroscience rapidly becoming a critical tool for understanding impression formation.

References

- Addis, D., Wong, A., & Schacter, D. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–1377.
- Addis, D., Wong, A., & Schacter, D. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science*, 19(1), 33-41.
- Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, 3(12), 469-479.
- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, 4(3), 165-178.
- Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1995). Fear and the human amygdala. *Journal of Neuroscience*, 15(9), 5879-5891.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268-277.
- Ballew, C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, 104(46), 17948-17953.
- Blakemore, S. J., Ouden, H., Choudhury, S., & Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions. *Social Cognitive and Affective Neuroscience*, 2(2), 130-139.
- Barresi, J. & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, 19(1), 107-122.
- Buckner, R. L., & Carroll, D. C. (2006). Self-projection and the brain. *Trends in Cognitive Sciences*, 11, 49-57.
- Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited:

- Universal structures plus writing system variation. *Human Brain Mapping*, 25(1), 92-104.
- Caramazza, A. (2000). The organization of conceptual knowledge in the brain. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 901-914). Cambridge, MA: MIT Press.
- Card, G., & Dickinson, M. H. (2008). Visually mediated motor planning in the escape response of drosophila. *Current Biology*, 18(17), 1300-1307.
- Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., et al. (2008). Cultural specificity in amygdala response to fear faces. *Journal of Cognitive Neuroscience*, 20(12), 2167-2174.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, 45, 3105–3113.
- Cohen, D., Vandello, J., Puente, S., & Rantilla, A. (1999). "When you call me that, smile!" how norms for politeness, interaction styles, and aggression work together in southern culture. *Social Psychology Quarterly*, 62(3), 257-275.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92(4), 631-48.
- Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science*, 19(2), 152-160.
- Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, 16, 1717-1729.
- Damasio, A. R. (1994). *Descartes' error*. New York: Putnam.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate

- the neural systems of reward during the trust game. *Nature Neuroscience*, 8, 1611-1618.
- Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, 84(6), 3072-3077
- Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature*, 8(3), 231-246.
- Duncan, J. & Owen, A. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Cognitive Sciences*, 23, 475-483).
- Eberhardt, J. L., & Fiske, S. T. (1998). *Confronting racism: The problem and the response*. Thousand Oaks, CA: Sage Publications.
- Engell, A., Haxby, J., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, 19(9), 1508-1519.
- Fiske, S. T., Cuddy, A., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77-83.
- Fiske, S. T. (2004). *Social beings: A core motives approach to social psychology*. Hoboken, NJ: Wiley.
- Fiske, S. T., & Taylor, S. E. (2008). *Social cognition: From brains to culture*. Boston: McGraw-Hill Higher Education.
- Fletcher, P. (1995). Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition*, 57(2), 109-128.
- Frith, C. D. (2007). The social brain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 671-678.
- Frith, C. D., & Frith, U. (2006a). How we predict what other people are going to do. *Brain*

- Research*, 1079, 36-46.
- Frith, C. D., & Frith, U. (2006b). The neural basis of mentalizing. *Neuron*, 50(4), 531-534.
- Frith, U., & Frith, C. D. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, 28(6), 151-155.
- Gallagher, H. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, 38(1), 11-21.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind". *Trends in Cognitive Sciences*, 7(2), 77-83.
- Gallagher, H., Jack, A., Roepstorff, A., & Frith, C. (2002). Imaging the intentional stance in a competitive game. *Neuroimage*. (16, 3 Part 1), 814-821.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T., Fiske, & G. Lindzey, (Eds.) *The handbook of social psychology* (4th ed.) (pp. 89-150). New York: McGraw Hill.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-21.
- Gilbert, D., Pelham, B., & Krull, D. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54(5), 733-740.
- Gilovich, T. (1987). Secondhand information and social judgment. *Journal of Experimental Social Psychology*, 23(1), 59-74.
- Gobbini, M., Koralek, A., Bryan, R., Montgomery, K., & Haxby, J. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, 19(11), 1803-1814.
- Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *Neuroreport*, 6(13), 1741-1746.
- Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait

- associations with faces. *Social Cognition*, 27, 222-248.
- Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes alzheimer's disease from healthy aging: Evidence from functional MRI. *Proceedings of the National Academy of Sciences*, 101(13), 4637-4642.
- Gutchess, A. H., Welsh, R. C., Boduroglu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing. *Cognitive, Affective, and Behavioral Neuroscience*, 6, 102-109.
- Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, 17(10), 847-853.
- Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience*, 2, 45-51.
- Harris, L. T., & Fiske, S. T. (in press). Dehumanized perception: The social neuroscience of thinking (or not thinking) about disgusting people. In M. Hewstone, & W. Stroebe (Eds.), *European review of social psychology*. London: Wiley.
- Harris, L. T., McClure, S. M., Van den Bos, W., Cohen, J. D., & Fiske, S. T. (2007). Regions of the MPFC differentially tuned to social and nonsocial affective evaluation. *Cognitive, Affective & Behavioral Neuroscience*, 7(4), 309-316.
- Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage*, 28(4), 763-769.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, 11(11), 2351-2351.
- Harvey, P. O., Fossati, P., & Lepage, M. (2007). Modulation of memory formation by stimulus content: Specific role of the medial prefrontal cortex in the successful encoding of social

- pictures. *Journal of Cognitive Neuroscience*, 19(2), 351-362.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664), 1634-1640.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-232.
- Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Eberhardt, J., Davies, P., Purdie-Vaughns, V., & Johnson, S., (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes, *Psychological Science*, 17(5), 383-386.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person-perception. *Advances in Experimental Social Psychology*, 2, 219-266.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1), 1-24.
- Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10(12), 1625-1633.
- Kampe, K., Frith, C. D., & Frith, U. (2003). "Hey John": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *Journal of Neuroscience*, 23(12), 5258-5263.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192—238). Lincoln: University of Nebraska Press.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*,

308(5718), 78-83.

Lieberman, M. D. (in press). Social Cognitive Neuroscience. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed.). New York: Wiley.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199-249.

Macrae, C. N., & Quadflieg, S. (in press). Person perception. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed.). New York: Wiley.

Mason, M. F., Banfield, J. F., & Macrae, C. N. (2004). Thinking about actions: The neural substrates of person knowledge. *Cerebral Cortex*, 14(2), 209-214.

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, 22(2), 171-193.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419-457.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserved person and object knowledge. *Proceedings of the National Academy of Sciences*, 99(23), 15238-15243.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, 24(21), 4912-4917.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex.

- NeuroImage*, 26(1), 251-257.
- Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience*, 29, 417-448.
- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, 27, 237-254.
- Oosterhof, N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087-11092.
- Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, 9, 128-133.
- Otten, L. J., & Rugg, M. D. (2001). Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex*, 11(12), 1150-1160.
- Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, 99(17), 11458-11463.
- Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, 57(1), 27-53.
- Phelps, E., O'Connor, K., Cunningham, W., Funayama, E.S., Gatenby et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation *Journal of Cognitive Neuroscience*. 12(5), 729-738.
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495-498.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546-550.

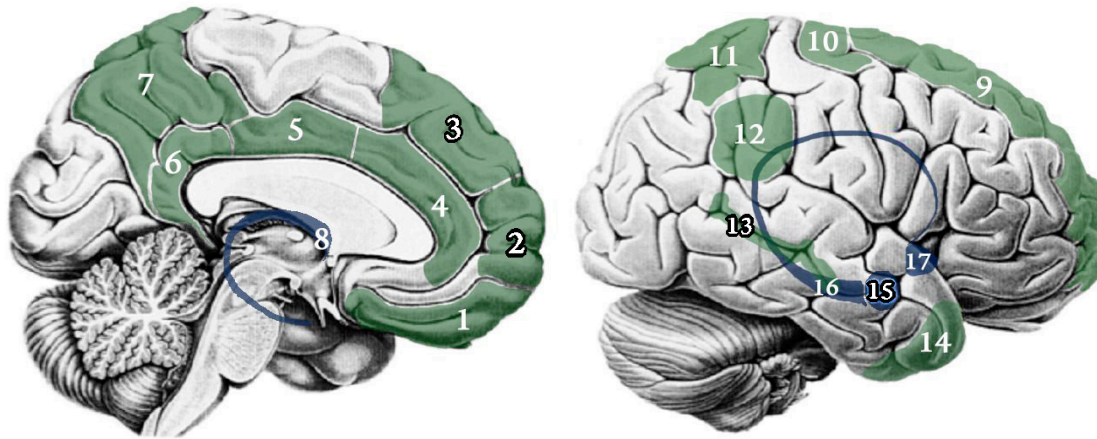
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22(4), 1694-1703.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 10), New York: Academic Press.
- Rossel, S., Corlija, J., & Schuster, S. (2002). Predicting three-dimensional target motion: How archer fish determine where to catch their dislodged prey. *Journal of Experimental Biology*, 205(21), 3321-3326.
- Said, C., Baron, S., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, 21(3), 519-528.
- Said, C., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9, 260-264.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: FMRI investigations of theory of mind. *NeuroImage*, 19, 1835-1842.
- Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, 17(8), 692-699.
- Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, 42(11), 1435-1446.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, 8(9), 657-661.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events:

- Concepts, data, and applications. *Annals of the New York Academy of Sciences*, 1124, 39-60.
- Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, 12, 508-514.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466-469.
- Singer, T., Winston, J., Kiebel, S., Dolan, R., & Frith, C. (2004). Brain responses to the acquired moral status of faces. *Neuron*, 41(4), 653-62.
- Squire, L. (1992). Memory and the hippocampus: A Synthesis From findings with rats, monkeys, and humans. *Psychological Review*, 99(2), 195-231.
- Taber, K., Wen, C., Khan, A., & Hurley, R. (2004). The limbic thalamus. *Journal of Neuropsychiatry and Clinical Neurosciences*, 16, 127-132.
- Thompson, M. S., Judd, C. M., & Park, B. (2000). The consequences of communicating social stereotypes. *Journal of Experimental Social Psychology*, 36(6), 567-599.
- Todorov, A. (in press). Evaluating faces on social dimensions. In A. Todorov, S. T. Fiske & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind*. Oxford University Press.
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, 3(2), 119-127.
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45(1), 163-173.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623-1626.

- Todorov, A., Pakrashi, M., & Oosterhof, N. (in press). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455-460.
- Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*. 83(5), 1051-65.
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39(6), 549-562.
- Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, 87, 482-493.
- Trope, Y., & Gaunt, R. (2003). Attribution and person perception. In M. A. Hogg, & J. Cooper (Eds.), *The sage handbook of social psychology* (pp. 190-209). New York: Sage Publications.
- Van den Bos, W., McClure, S. M., Harris, L. T., Fiske, S. T., & Cohen, J. D. (2008). Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex. *Cognitive, Affective and Behavioral Neuroscience*, 7(4), 337-46.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829-58.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention *Trends in Cognitive Sciences*, 9(12), 585-594.
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., & L. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, 16(10), 1854-1863.

- Winston, J. O'Doherty J., Dolan R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, 20(1), 84-97.
- Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, 40(4), 1912-1920.

Figure 1



Medial Surface

Lateral Surface

- (1) Orbitofrontal Cortex
- (2) Ventral Medial Prefrontal Cortex**
- (3) Dorsal Medial Prefrontal Cortex**
- (4) Anterior Cingulate Cortex
- (5) Middle Cingulate Cortex
- (6) Posterior Cingulate Cortex
- (7) Medial Parietal Cortex
- (8) Caudate Nucleus (not actually visible)

- (9) Superior Frontal Gyrus
- (10) Superior Precentral Gyrus
- (11) Superior Parietal Gyrus (Superior Parietal Lobule)
- (12) Temporoparietal Junction
- (13) Superior Temporal Sulcus**
- (14) Temporal Poles
- (15) Amygdala** (not actually visible)
- (16) Hippocampus (not actually visible)
- (17) Anterior Insula (not actually visible)

Figure 2

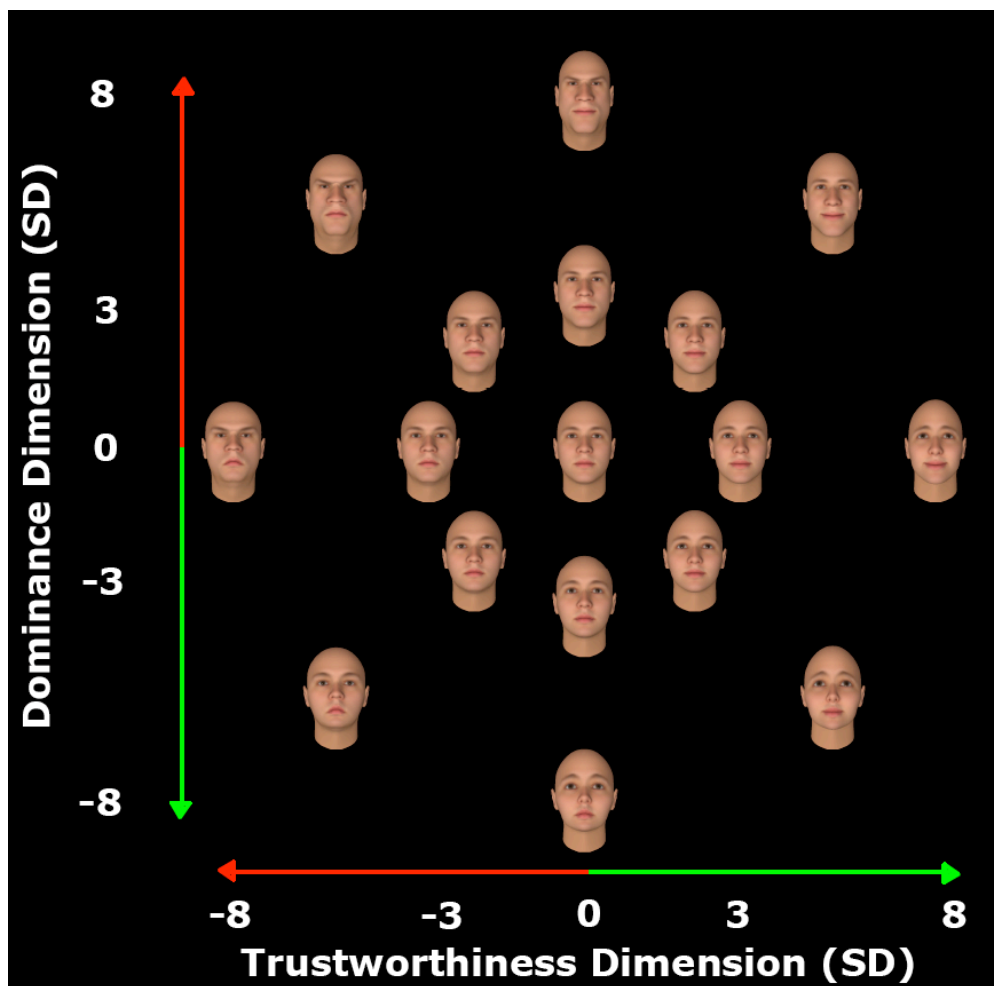


Figure 3

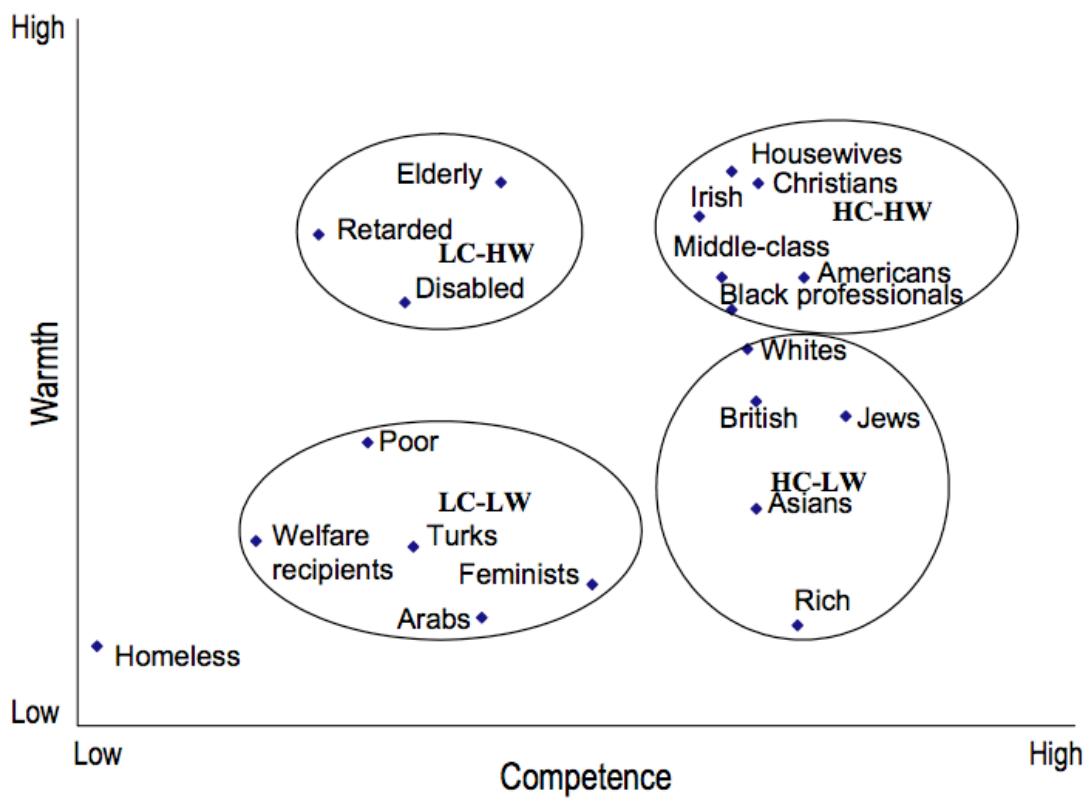


Figure Captions

Figure 1. Brain regions featured in this chapter. Those regions that are discussed most prominently are labeled with black-outlined numbers and are bolded in the figure key. Darker shades of gray are assigned to structures located *between* the lateral (side) and medial (center-line) surfaces, but which are displayed *on* the lateral or medial surface for presentation purposes (caudate, hippocampus, amygdala, anterior insula). Adapted from Lieberman (in press).

Figure 2. A data-driven model of face evaluation. Human judges rated 300 emotionally neutral faces on the primary social dimensions of dominance and trustworthiness. The appearance of each face was statistically represented in terms of 50 independent principal components. Covariation between these principal components and participants' ratings revealed what aspects of facial physiognomy most strongly drive people's judgments of facial trustworthiness and dominance. Modeling extremely trustworthy/untrustworthy and dominant/submissive faces (here, up to 8 standard deviations from the mean on each dimension) strongly suggests that trustworthiness judgments derive from an overgeneralization of emotional expressions of happiness and anger. Perceptions of dominance (orthogonal to trustworthiness) correlate with perceived masculinity/femininity.

Figure 3. Scatter plot and cluster analysis of competence and warmth ratings for various social groups (reproduced, with permission, from Cuddy et al., 2007). The Stereotype Content Model proposes that social groups are readily perceived in terms of two orthogonal dimensions: warmth and competence. Participants (in this case, US

survey respondents) rated each group on warmth and competence using a 5-point scale. Ratings were then submitted to cluster analysis, yielding the four circled clusters shown here. Groups appearing near the center of clusters most reliably replicate their cluster membership across studies.