CHAPTER 19

Automaticity, Control, and the Social Brain

Robert P. Spunt and Matthew D. Lieberman

The social world is good at keeping the human brain busy, posing cognitive challenges that are complex, frequent, and enormously important to our well-being. In fact, the computational demands of the social world may be the principal reason why the human brain has evolved to its present form and function relative to other primates (Dunbar, 1993). Importantly, the human brain is often able to make sense of the social world without having to do too much work. This is because many of its processes are automatically initiated by the presence of relevant social stimuli and run to completion without much, if any, conscious intervention (Bargh & Chartrand, 1999; Gilbert, Pelham, & Krull, 1988). Not surprisingly, then, the idea that mental processes vary in their level of automaticity has had a powerful influence on theories of social cognition (Chaiken & Trope, 1999).

In the past two decades, the rapid development of methods for peering into the working human brain has made it possible to study the operation of the brain as it confronts and reflects on social information. As a result, we now know a great deal about the neuroanatomical correlates of social cognition. Although automaticity and control have been invoked to describe the operation of the social brain (cf. Lieberman, Gaunt, Gilbert, & Trope, 2002; Lieberman, 2007; Morris & Mason, 2009), there appears to be no consensus regarding what these constructs mean in terms of brain function or how they should be investigated using the methods of social neuroscience (but see Satpute & Lieberman, 2006). Hence, we begin this chapter by broadly considering how the construct of automaticity should be understood and investigated in social neuroscience. Next, we narrow our focus to social neuroscience research, to a fundamental topic in social cognition, in which questions of automaticity have historically had a large influence: causal attribution of observed behavior. Such research has identified two large-scale systems in the human brain that are reliably associated with dissociable functions in the causal attribution of behavior: the so-called mirror neuron and mentalizing systems. Now that these systems have been identified, social neuroscience has reached a point that questions about the automaticity of social causal attribution can be tackled by investigating the operating characteristics of the brain systems with which that process is reliably associated. Based on existing research, we propose an identification–attribution (I-A) model of these two systems' contributions to the causal explanation of behavior, with the mirror neuron system supporting relatively automatic behavior identification and the mentalizing system supporting relatively controlled social causal inference. After introducing the model, we evaluate existing evidence regarding the operating characteristics of these two systems. Finally, we consider directions for future research.

NEURAL AUTOMATICITY

What do the constructs of automaticity and control mean in terms of the social brain? Figure 19.1 depicts what could be called the *categorical framework*, so named because it treats automaticity and control as separate processing categories that can each be described by a set of operating characteristics. *Automatic processes* are those that operate unconsciously, efficiently, unintentionally, and in a manner incapable of being controlled, while *controlled processes* are those that do not possess these characteristics. The position of the arrows on the brain image in Figure 19.1 illustrates the conflation of bottom-up (i.e., stimulus-driven) processing in the brain with automaticity, and top-down (i.e., goal-driven) processing in the brain with control.

The categorical framework, although intuitive and in many cases useful, glosses over a great deal of complexity. Here, we follow Bargh (1989, 1994) in advocating a multidimensional framework in which the automaticity profile of a neurocognitive process is described in terms of a set of partially orthogonal dimensions. These dimensions are *awareness*, *efficiency*, *intentionality*, and *controllability*. In the first part of this section, we define these dimensions and identify methods for investigating them in the brain (Table 19.1). Following this, we conclude the section by proposing that claims regarding neural automaticity should include a specification of the conditions of operation.

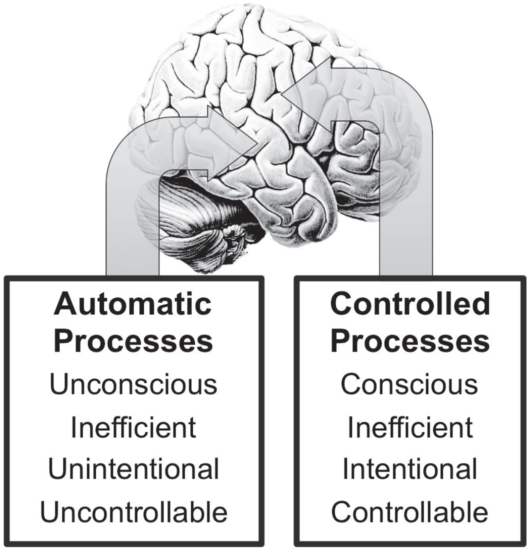

FIGURE 19.1. The categorical view of automaticity and control in the brain.

Awareness

Awareness, often used synonymously with *consciousness,* refers to the extent to which the subject is phenomenally aware of the stimulus that initiates a process, the operation of the process itself, and/or the output of the process. What does it mean to say that a person has awareness of a neural process? In the current state of the psychological and brain sciences, this is more a metaphysical question than a scientific one. However, from a methodological perspective one can establish operation *without* awareness by testing whether subliminal stimulus variation impacts neural activity. In terms of experimental design, the logic is identical to behavioral studies employing subliminal primes (Bargh & Chartrand, 2000). Yet unlike behavioral studies, the researcher is not dependent on measuring a behavioral outcome (e.g., response latency, recall) to establish that the subliminal stimulus induced a nonconscious process. Instead, measures of brain activity can be used to establish nonconscious processing. For example, numerous functional magnetic resonance imaging (fMRI) studies have now observed amygdala sensitivity to variation in the emotional expression (Whalen et al., 1998) or value (Morris, Ohman, & Dolan, 1999) of faces presented subliminally. On the basis of such data, one can conclude that some component of face processing in the amygdala *can* occur in the absence of awareness. However, this does not permit the conclusion that face processing in the amygdala is *always* automatic or that other functions associated with the amygdala are automatic (more on this later).

TABLE 19.1. Four Dimensions of Neural Automaticity: Definitions and Methods of Investigation

Dimension	Definition	Methods
Awareness	The extent to which the subject is phenomenally aware of the stimulus, process, and/or response	1. Subliminal stimulus variation (e.g., Whalen et al., 1998) 2. Induce and measure implicit versus explicit learning (e.g., Foerde et al., 2006)
Efficiency	The extent to which a process can run to completion quickly and in the absence of effort and/or attention	1. Dual-task paradigms (e.g., Poldrack et al., 2005) 2. Task-irrelevant stimulus variation (e.g., Winston et al., 2002) 3. High-resolution measurement of response timing (e.g., Van Berkum et al., 2009)
Intention	The extent to which a process is initiated by an explicit intention to do so	1. Manipulate processing goal (e.g., Hesse et al., 2008) 2. Manipulate processing demand (i.e., difficulty/load) (e.g., McKiernan et al., 2003)
Control	The extent to which an ongoing process can be altered by an explicit intention to do so	Initiate process of interest, then induce goal to alter processing (e.g., Ochsner et al., 2002)

Another method for tapping process awareness involves establishing a disconnect between self-reported behavioral intentions and a neural process known to be associated with the intended behavior. To the extent that such a disconnect is observed, one can conclude that the subject is unaware of the neural process. Using this logic, researchers have suggested that the formation of movement intentions operates unconsciously insofar as neural activity associated with the production of movements actually precedes participants' self-reported decision to move (Desmurget & Sirigu, 2009). Using a similar logic, work from our group has shown in several studies that neural processes at work during the consumption of persuasive messages do a better job of predicting participants' ensuing behavior than do participants' own self-reported intentions to engage in those behaviors (Falk, Berkman, Mann, Harrison, & Lieberman, 2010; Falk, Berkman, Whalen, & Lieberman, 2011). On the flip side, claims that a neural process is available to awareness can be supported by observations of tight coupling between self-reported phenomenal states (e.g., frustration) and a neural process (e.g., the neural response to errors; Spunt, Lieberman, Cohen, & Eisenberger, 2012). These studies suggest that in addition to the use of subliminal stimulus variation, the careful assessment of (self-reported) phenomenal experience can be used to make claims about processing (un)awareness.

Efficiency

The *efficiency of a neurocognitive process* can be defined as the extent to which it can be executed quickly and in the absence of attention. Phenomenologically, efficient processes are fast and effortless (e.g., recognizing a familiar face), whereas less efficient processes are slow and effortful (e.g., mentally computing the product of 42 and 79). Importantly, efficient processing is not the same as *spontaneous* processing. Spontaneity simply means that a process will be engaged even in the absence of an external stimulus to do so (e.g., the instruction to multiply 42 and 79). Hence, a process that begins spontaneously need not operate efficiently.

The principal method for studying processing efficiency is the dual-task paradigm, in which the researcher gives the subject a primary task known to engage the process of interest and simultaneously varies either the presence or difficulty of a secondary task (often called a manipulation of "cognitive load"). When a dual-task paradigm is employed in a behavioral study, questions about efficiency can only be answered if both tasks produce measurable behavioral outcomes that validly reflect successful engagement of the primary and secondary processes. When such outcomes are available, one can confirm that cognitive load was successfully induced (by examining performance of the secondary task) and assess the extent to which it affects performance on the primary task. If performance on the primary task is unaffected by load, one can infer that the process associated with the primary task is efficient.

Dual-task paradigms can be used to study neural efficiency by engaging the neural process of interest in a primary task, while simultaneously varying either the presence or difficulty of a secondary task. However, when measuring brain activity, one is no longer dependent on observations of performance variability, since variability in the neural response is now a (more) direct measure of the process of interest. Hence, primary tasks that do not produce a behavioral outcome, such as supraliminal stimulus manipulations, can be used to study neural efficiency (e.g., Pessoa, McKenna, Gutierrez, & Ungerleider, 2002). However, it should be emphasized that in order for a supraliminal stimulus manipulation to provide evidence of neural efficiency, it must be paired with a cognitively demanding secondary task, so that strategic stimulus processing can be ruled out as an alternative explanation. Moreover, we note that to categorize such a paradigm as "dual-task" is perhaps a misnomer, since passive stimulus perception is not a "task" in the traditional sense. Hence, it may be more appropriate to characterize this type paradigm as involving "task-independent stimulus variation."

Using dual-task paradigms in neuroimaging experiments is not without its problems (Satpute & Lieberman, 2006). Perhaps the most significant problem occurs when the primary and secondary tasks have overlapping neural correlates. When they do, a region whose level of activity is unaffected by the introduction of the secondary task may demonstrate such an effect not because of efficient operation in the primary task but because of its involvement in the secondary task. Hence, it is important in studies of neural efficiency to establish that regions demonstrating evidence of efficient operation are not associated with performance of the secondary task (e.g., in an independent localizer scan and/or by reference to meta-analytic evidence). Moreover, while it is the case that neural efficiency can be studied using the neural response alone as the outcome of interest, having behavioral indices of efficient operation (e.g., accuracy, response latency) can help constrain interpretation of such effects (as in Foerde, Knowlton, & Poldrack, 2006). For instance, a region that responds with equal magnitude to a particular stimulus category (e.g., faces) regardless of level of load is consistent with efficient processing. However, if its response is correlated with some behavioral outcome (e.g., postscan recall) only in conditions of low load, then this can be used to infer inefficient processing.

Processing efficiency is often defined in terms of the speed with which the process can be completed. Studies using fMRI are somewhat ill equipped to study processing speed, because the measured signal is dependent on the relatively slow hemodynamic response to neural activity, hence offering temporal resolutions on the order of seconds. Instead, researchers interested in the speed of the neural response must turn to techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), which provide temporal resolution on the order of milliseconds. Using such techniques, researchers have, for instance, shed light on efficient processing of action-related language in the brain's motor system (for a review, see Hauk, Shtyrov, & Pulvermüller, 2008).

Intentionality and Controllability

The dimensions of intentionality and controllability are closely linked and are discussed together here. Whereas the *intentionality* of a mental process regards the extent to which its initiation depends on the presence of an explicit intention to initiate it, *controllability* refers to the extent to which the course of a process can be altered (i.e., modulated or terminated) after it has begun. Methodologically, we consider two ways in which the intentionality of a region's operation can be investigated. The first involves manipulating the subject's goal while presenting no stimulation. Using such a method, researchers have, for example, demonstrating that visual cortices of the brain are subject to intentional control insofar as they activate when subjects intentionally engage in visual imagery (for a review, see Kossyln, Ganis, & Thompson, 2001). A second method involves manipulating the subject's goal while either holding stimulation constant (e.g., Spunt, Falk, & Lieberman, 2010) or crossing the goal manipulation with a stimulus manipulation (e.g., Winston, Strange, O'Doherty, & Dolan, 2002; Spunt & Lieberman, 2012b).

To investigate controllability, the researcher must first induce the process of interest (e.g., a negative emotional response to an aversive stimulus), then manipulate the presence of a goal to in some way alter the process (e.g., diminish the negative emotional response by reappraising the stimulus; Ochsner, Bunge, Gross, & Gabrieli, 2002). This method allows the researcher to assess the dynamic interaction of bottom-up (i.e., stimulus-driven) and top-down (i.e., goal-driven) psychological processes. In order to do so, studies employing this method typically proceed by identifying a region (or set of regions) as implementing the top-down process (e.g., searching the brain for regions associated with the presence and/or successful execution of the control goal) and a region (or set of regions) whose association with the bottom-up process is affected by the presence and/or successful execution of the control goal. Furthermore, these studies can employ a variety of connectivity analyses to determine whether activation of the region identified as "controller" exhibits its dynamic changes in activity that track goal-dependent changes in activation of the region(s) identified as "target" (Lieberman, 2011).

When the process to be controlled is initiated by presentation of a stimulus, researchers should be careful in interpreting goal-dependent changes in neural processing of the stimulus (Pessoa, Kastner, & Ungerleider, 2003). This is because these changes could be caused by a direct influence of the controller region(s) on the target region(s) or by an indirect path in which the controller region(s) produce shifts in spatial attention that ultimately alter the bottom-up inputs to target regions (Posner, 1980). This issue can be at least partially addressed by tracking eye movements (e.g., Dalton et al., 2005). Yet even when spatial attention is confirmed by eye-tracking data, it remains possible that the controller has an influence not by directly altering the target but by altering inputs to the target.

Neural Automaticity Is Conditional

It is important to emphasize that claims about automaticity should regard specific functions of regions rather than the regions themselves. For instance, consider the case of the amygdala, often invoked as a paradigmatic example of neural automaticity. Although some studies have shown that this region responds to subliminally presented threat cues (e.g., Morris et al., 1999), other work suggests that the automatic response of the amygdala to such cues is dependent on the presence of sufficient attention to the sensory modality in which such cues are presented (e.g., Pessoa et al., 2002). For example, Mothes-Lasch, Mentzel, Miltner, and Straube (2011) simultaneously presented threatening auditory cues (angry or neutral voices) and neutral visual stimuli (cross or circle) to subjects and manipulated whether they judged the gender of the speaker or the type of visual symbol. Replicating previous research, the amygdala robustly responded when participants attended to the voice during the gender judgment task. However, this effect was extinguished when attention was instead directed toward the visual modality in the symbol judgment task. Studies such as this one suggest a conditional view of automaticity in the amygdala that be phrased as *When a threat cue is present in a sensory modality to which attention is being directed, the amygdala can detect it even in the absence of explicit awareness*. In addition, such statements might also include the method used to measure the amygdala response insofar as different methods capture the response at different timescales (cf. Luo et al., 2010).

Thus, in addition to the multidimensional approach presented earlier, it is important to recognize that neural automaticity is *conditional* (Bargh, 1989). This is because regions do not operate in isolation but are part and parcel of distributed functional networks that are constantly being modulated by changes in the internal and external environment (Pessoa, 2008). Even holding the region constant, function can vary dramatically depending on the conditions under which it is engaged (Poldrack, 2006). In addition, it is likely the case that even when holding the region and function constant, the automaticity profile may be conditional. These considerations must be kept in mind when using the construct of automaticity to talk about and study the working brain.

Of course, we note that anatomy-based variables can be useful for informing theories about the operating characteristics of neural processes. One such variable is the distance from the sensory periphery. This concept is helpful insofar as regions that are closer to the sensory periphery (e.g., visual cortices) are probably more likely to exhibit high levels of automaticity than are regions more distant from the sensory periphery (e.g., prefrontal areas). Another such variable is the connectivity profile of a brain region. To the extent that a region has extensive connectivity with the rest of the brain, it is more likely to operate in different ways depending on conditions of its operation (Pessoa, 2008).

BRAIN SYSTEMS FOR THE CAUSAL EXPLANATION OF BEHAVIOR

Given the many demands placed on the brain by the social world, our discussion of the existing social neuroscience research pertaining to automaticity is limited to just one of its demands: explaining the behavior of others. The causal explanation of behavior has a long history in social psychology, studied primarily under the name of attribution theory (Heider, 1958; Jones & Harris, 1965). Moreover, the distinction between automaticity and control has had a strong influence on attribution theories, producing several dual-process models of the attributional inference (Gilbert, 1989; Trope, 1986; Lieberman et al., 2002). Although there are notable differences among existing models, they generally agree that attributional inferences are the result of an initial stage in which stimulus input about behavior is automatically identified, and a subsequent stage in which the identified behavior is attributed to an inferred social cause, using processes that are at least partially dependent on control. Importantly, these models are of the dual-process ilk in two ways. One, they specify two dissociable processes: identification and attribution. Two, they specify two dissociable operating characteristics of these processes: automaticity and control.[1] In this section, we start with the former, which predicts that identification and attributional processes will rely on dissociable neural processes.

What does it mean for an observer to identify another person's behavior? From the perspective of attribution theory, the output of the identification process must produce a representation that is attribution-relevant, that is, capable of being explained in terms of the actor's underlying disposition (Jones & Harris, 1965). From the perspective of perception, behavior identification entails extracting from the complex and ever-changing inflow of sensory input an understanding of *what* it is that the other person is doing with his or her body, that is, his or her motor intention. Social neuroscience suggests that in order for an observer to understand what *another person* is doing with his or her body, the observer may use some of the same processes involved in using *his or her own* body. Not surprisingly, a distributed set of regions in the human brain reliably responds when individuals produce motor actions such as grasping. What may be more surprising is that many of these same regions are active when individuals merely watch other people producing similar motor actions. This set of regions has come to be known as the human *mirror neuron system* (MNS; Figure 19.2A; Rizzolatti & Craighero, 2004), and canonically includes dorsal and ventral regions of the premotor cortex (dPMC and vPMC) and a region of the parietal cortex in the rostral inferior parietal lobule (IPL) extending into the intraparietal sulcus (IPS).[2] The concept of a mirror neuron is based on studies of single-cell activity in the macaque brain, which have mapped neurons responding similarly to the production and perception of motor actions. To summarize this rather large and influential body of animal research, it appears that collectively these "mirror neurons" evidence a rapid neural process of translating sensory input about another person's motor action (e.g., the sight of a hand moving toward a cup of coffee) into a representation of the action's goal (the person is going to grip the cup; Gallese, Fadiga, Fogassi, & Rizzolatti,

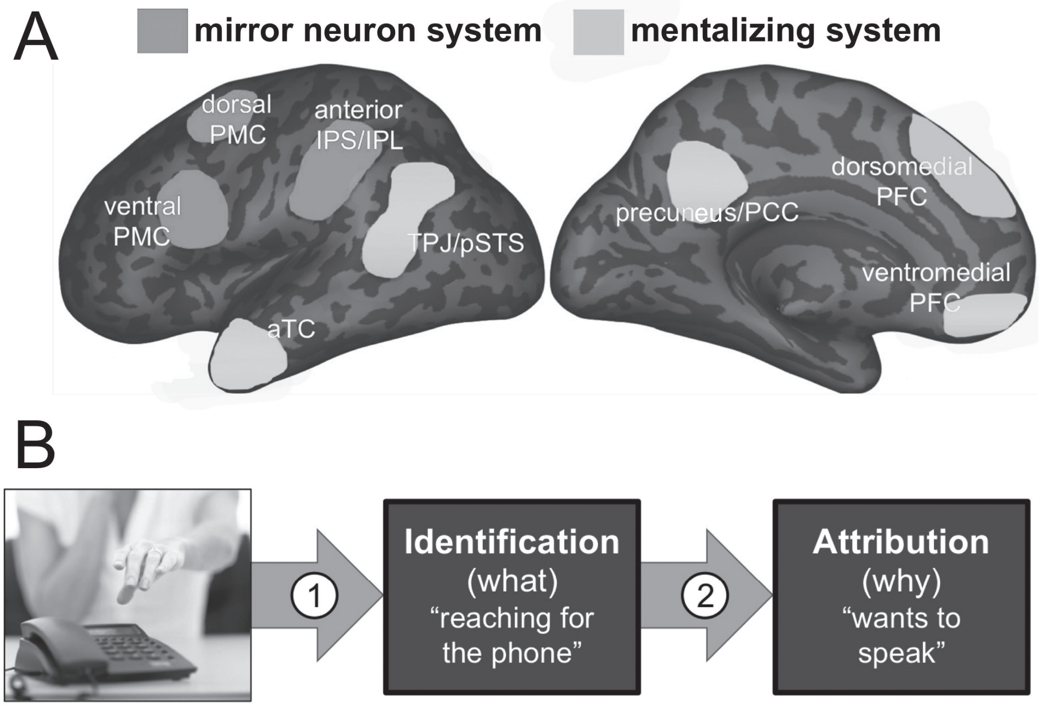

FIGURE 19.2. (A) Two brain systems involved in social cognition. PMC, premotor cortex; IPS, intraparietal sulcus; IPL, inferior parietal lobule; aTC, anterior temporal cortex; TPJ, temporoparietal junction; pSTS, posterior superior temporal sulcus; PCC, posterior cingulate cortex; PFC, prefrontal cortex. (B) The identification–attribution model of MNS and MZS contributions to understanding observed behaviors. The MNS is hypothesized to contribute primarily to identifying observed motor behaviors (Path 1), while the MZS is hypothesized to contribute primarily to attributing identified behaviors to inferred social causes (Path 2).

1996) and the intended outcome of reaching that goal (the person is going to raise the cup to his or her mouth; Fogassi et al., 2005). Neuroimaging studies suggest that similar processes are at work in the human MNS (Zacks et al., 2001; Iacoboni et al., 2005; Hamilton & Grafton, 2006, 2008; de Lange, Spronk, Willems, Toni, & Bekkering, 2008).

Imagine seeing a man reach for glass of scotch at 10:00 A.M. Your friend asks you, "Why is he doing it?" and you answer, "In order to take a drink." Your rather unsatisfactory conclusion is a verbalization of the hypothesized contribution of the MNS to action understanding (Iacoboni et al., 2005; Gallese, 2007). Although this answer is technically and grammatically correct, it is most assuredly not the kind of answer your friend was looking for. If Fritz Heider were alive today, he might say that the MNS ends at the beginning of an inferential process that concludes not with observable actions but with unobservable entities such as motive, belief, character, and context. Inferring these kinds of entities—mental states, traits, and social context—is referred to as an act of *mentalizing* (Frith & Frith, 2006). Within social neuroscience, the discovery of mirror neurons is rivaled in importance by the discovery of a collection of regions in the human brain that reliably activate when individuals are prompted to think about the mental states of others (Frith & Frith, 2006; Saxe, 2006). This *mentalizing system* (MZS; Figure 19.2A) canonically includes the dorsomedial prefrontal cortex (dmPFC), an area of the medial parietal cortex spanning the precuneus and posterior cingulate cortex (PC/PCC), the temporoparietal junction at the posterior end of the superior temporal sulcus (TPJ/pSTS), and the anterior temporal cortex (aTC). Although the earliest work on the MZS primarily relied on either verbal descriptions or abstract depictions of human behavior, more recent work has demonstrated that the MZS does come online to make sense of observed actions (Grèzes, Frith, & Passingham, 2004; Brass, Schmitt, Spengler, & Gergely, 2007; de Lange et al., 2008).

On the basis of this prior work, we have tested an I-A model of MNS and MZS contributions to social cognition, in which the MNS supports the identification of observed motor behaviors (Path 1 in Figure 19.2B), while the MZS supports the process of inferring the social causes that explain identified behaviors (Path 2 in Figure 19.2B). Put another way, the MNS makes sense of what a body is doing so that the MZS can use that information to make sense of what is going on in the person's mind. To explicitly investigate the I-A model we have developed a general method called the *why/how paradigm* (Spunt et al., 2010; Spunt, Satpute, & Lieberman, 2011). The paradigm is simple: Participants are shown a behavioral stimulus (e.g., a video clip or description of a familiar, goal-directed action) and on different trials are induced to think of one important part of *how* the behavior is occurring or one plausible reason explaining *why* the behavior is occurring. This is a face-valid manipulation of identification and attribution goals, and features high ecological validity, using natural language to induce spontaneous, open-ended social cognition. Given that the task is a goal manipulation, it permits using the same stimuli in both conditions. This feature makes it feasible to use highly complex and naturalistic social stimuli, and offers the possibility of teasing part bottom-up stimulus effects from top-down goal-driven effects.

A recent study (Spunt & Lieberman, 2012a; Figure 19.3) illustrates well the use of the *why/how paradigm* to test the I-A model. While undergoing fMRI, healthy volunteers were presented with familiar human actions either by video or text and were asked either to identify *how* (identification goal) or to infer *why* (attribution goal). Identification and attribution goals strongly distinguished activity in the MNS and MZS, respectively, and these effects were present regardless of whether the behavior was presented in a video or text. However, some areas of the MNS showed a strong preference for behaviors presented in a video as opposed to in text, and this effect was present regardless of the observer's explicit goal. This suggests that the MNS may translate sensory information about behavior into meaningful motor acts, even when the observer's explicit goal is not to identify motor acts but to make causal attributions. Moreover, it suggests that the MNS may not be critically involved in making attributional inferences when such inferences are made from behaviors described in text. This is because, when actions are *observed*, sensory input about action must be identified in meaningful ways in order for attributional processing to begin. In support of this, we found that areas of the MNS and MZS were functionally coupled, but only when participants made attributions about *observed* behaviors. In contrast, when attributions were made for verbal descriptions of behaviors, such a functional coupling was not observed, a finding consistent with the proposition that only during action *perception* does the MZS depend on the MNS for the identification of behavior. This study, along with others from our group (Spunt et al., 2011; Spunt & Lieberman, 2012b), provides strong support for the I-A model of the functions of the MNS and MZS in social inference.

AUTOMATICITY PROFILES OF THE MNS AND MZS

If we watch someone perform a familiar action, it is as if we literally *see* what they are doing in a psychological sense. When we see someone reaching for a glass of water, the characterization as "reaching" comes to us as a given. We do not experience the act in terms of its complex physical kinematics. By the time we notice it, it is already infused with psychological interpretation. This psychological perception does not feel like it is based on any sort of information processing at all; it is, in a sense, given to us prior to any sort of cognitive elaboration. Of course, on the basis of such perceptions we can explicitly verbalize an understanding of the person's likely state of mind, such as his or her motives, beliefs, and personal character. Correspondent with this phenomenological

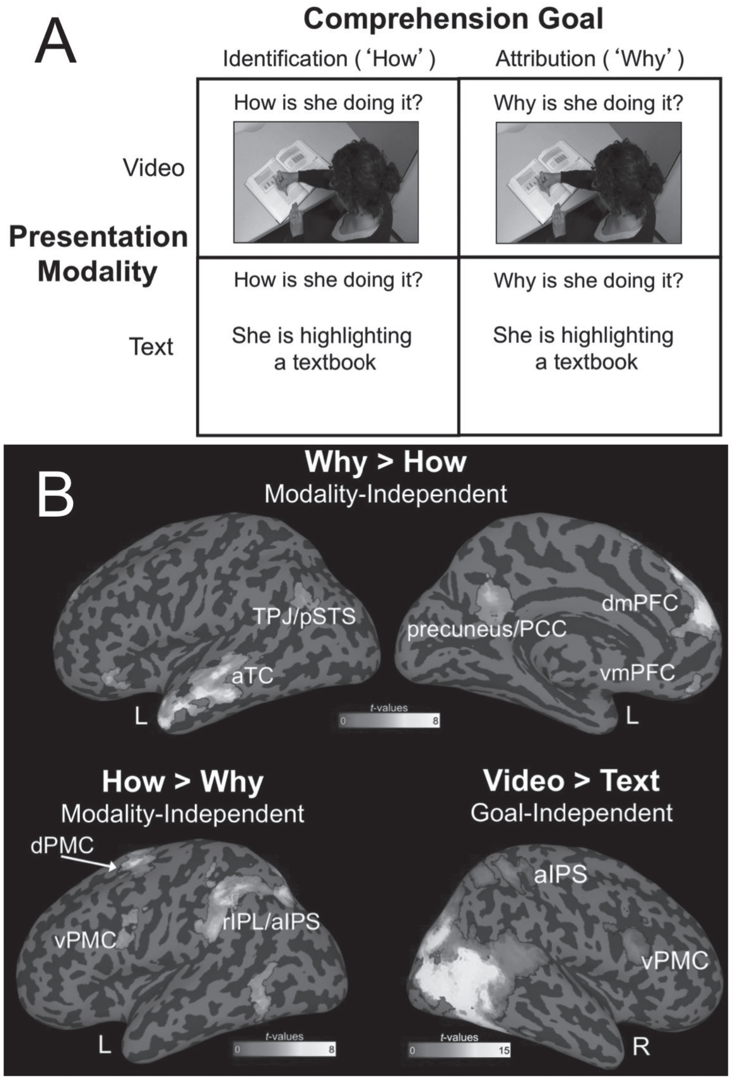

FIGURE 19.3. (A) The experimental design used in Spunt and Lieberman (2012a) and discussed in the text. (B) Regions recruited by the identification and attribution goals regardless of presentation modality, and regions recruited by the presence of an action in a video (compared to an action described in text) regardless of the observer's explicit comprehension goal. vPMC, ventral premotor cortex; dPMC, dorsal premotor cortex; aIPS, anterior intraparietal sulcus; rIPL, rostral inferior parietal lobule; aTC, anterior temporal cortex; TPJ, temporoparietal junction; pSTS, posterior superior temporal sulcus; PCC, posterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex.

description, it is tempting to assume that that the MNS operates automatically, while the MZS operates under relatively more control. In the previous section, we presented evidence associating the MNS with the *identification* of perceived behavior and the MZS with the *attribution* of such perceptions to inferred causes. In this section, we evaluate the case for automaticity and control in the MNS and the MZS, respectively. As we earlier tried to make clear, the categorical view of automaticity and control ought to be supplanted by a multidimensional view that describes neurocognition not as either automatic or controlled but as having a potentially complex and conditional automaticity profile. Therefore, we review existing social neuroscience evidence regarding the operating characteristics of the MNS and MZS in order to make some preliminary proposals regarding their automaticity profiles. We make no claim to comprehensiveness, and in order to narrow our scope we primarily focus on studies using fMRI.

Theoretical treatments of the MNS have often included the proposition that the system operates automatically (Gallese, 2003, 2007, 2009; Coricelli, 2005; Iacoboni et al., 2005; Keysers & Gazzola, 2006). However, this proposition has received very little empirical attention. To our knowledge, no study has explicitly examined the operating characteristics of mirror neurons in the macaque brain. In that animal research, the best evidence that might be marshaled for the automaticity claim seems to be the rapidity with which these neurons respond during action observation. However, the speed of response is not typically analyzed; moreover, speed is only one part of the *efficiency* dimension of automaticity. To establish efficiency, it would be necessary to load the monkey with a secondary task, while simultaneously presenting actions. Hence, we believe that on the basis of the monkey studies alone, there is only limited evidence for automaticity, and such evidence applies to only one dimension (efficiency) of automaticity. Moreover, it is unknown to what extent the efficiency of this response is dependent on explicit attention to the action stimulus. Insofar as monkeys require extensive training in order to participate in these experiments, it seems likely that appropriate attentional control is a condition of the efficient response of macaque mirror neurons.

In humans, the automaticity claim has received relatively more attention (selected studies are listed in Table 19.2). In an influential study attempting to investigate this claim, Iacoboni and colleagues (2005) had participants watch actions that either did or did not feature a context that suggested the action's intention. Some participants were also given the explicit goal to infer the action's intention, while others were simply instructed to watch passively. Contextualized actions elicited increased activation in the right vPMC, and this effect did not interact with the goal manipulation. This finding was used to support the claim that the context sensitivity in the right vPMC is automatic. In terms of the multidimensional framework articulated earlier (and summarized in Table 19.1), this claim to automaticity seems to primarily regard the *efficiency* with which the right vPMC encodes the stimulus manipulation. However, in order for a stimulus manipulation to provide evidence of efficient operation, it must be paired with a manipulation of attentional load that is unrelated to the stimulus manipulation. The explicit goal to infer intention, insofar as it is directly (and purposefully) related to the stimulus manipulation, is not an appropriate induction of attentional load. Therefore, though the result from Iacoboni et al. (2005) is consistent with efficiency operation, it does not demonstrate it and in fact did not even test it.

Two studies have used dual-task paradigms explicitly to investigate the efficiency of the MNS response during action observation. Chong, Williams, Cunnington, and Mattingley (2008) presented videos of moving hands, stationary hands, or random dot motion to participants and simultaneously manipulated the difficulty of a secondary visual discrimination task (visual load) that was superimposed onto the videos. Increasing visual load attenuated the response of left posterior inferior frontal gyrus (pIFG; pars triangularis) only during hand observation, a finding consistent with nonefficient operation of this region during action observation. However, it is useful to note that this region of the IFG is more anterior than is typically observed in studies of action observation and imitation (for a meta-analysis, see Caspers, Zilles, Laird, & Eickhoff, 2010). In a recent study (Spunt & Lieberman, 2013), we simultaneously manipulated participants' comprehension goal and their level of memory load while they watched videos of goal-directed actions. Compared to a fixation baseline, action observation elicited activation in bilateral vPMC, left dPMC, and left anterior IPS that was not unaffected by level of memory load. To our knowledge, this is the most direct evidence yet for an efficient response of the MNS to actions.

Although Iacoboni and colleagues (2005) did not find support for the controllability claim, several other studies have. Indeed, years earlier, Iacoboni and colleagues (1999) showed that the goal to imitate rather than

TABLE 19.2. Selected fMRI Studies Germane to the Automaticity Profile of the Mirror Neuron System (MNS) Response during Action Observation

Study	Method	Findings	Implications
Iacoboni et al. (1999)	Manipulated presence of goal to imitate observed actions	Imitation goal increased activation in left pIFG and aIPS	Some MNS areas are controllable
Iacoboni et al. (2005)	Manipulated (1) presence of context in observed actions and (2) presence of goal to infer action intention	Presence of context increased activation in right vPMC regardless of observer goal	Context senstivity in some MNS areas is spontaneous
Chong et al. (2008)	Manipulated visual load during observation of moving hands, stationary hands, or dot motion	Load decreased left pIFG activation only during hand observation	Some MNS areas operate nonefficiently under visual load
Hesse et al. (2008)	Manipulated attention to the means or the end of observed actions	Attention to means (relative to end) increased activation in bilateral aIPS/rIPL and right vPMC/pIFG	Some MNS areas are controllable
Spunt et al. (2011)	Manipulated presence of goal to verbally identify observed actions	Relative to passive observation, active identification increased activation in right pIFG and rIPL and left aIPS	Some MNS areas are controllable
Spunt & Lieberman (2012b)	Manipulated (1) attention to how or why actions are performed and (2) presentation modality (text vs. video)	Attention to how increased activation in left vPMC, dPMC, and aIPS/rIPL regardless of presentation modality	Some MNS areas are controllable and can be intentionally operated
Spunt & Lieberman (2013)	Manipulated (1) processing goal and (2) memory load during action observation	Load did not affect activation of bilateral vPMC and left dPMC and aIPS for any processing goal	Some MNS areas operate efficiently under memory load

Note. aIPS, anterior intraparietal sulcus; dPMC/vPMC, dorsal/ventral premotor cortex; dPMS; pIFG, posterior inferior frontal gyrus; rIPL, rostral inferior parietal lobule.

passively observe actions enhanced activation in the MNS. Later on, Hesse, Sparing, and Fink (2008) presented videos of hand actions while participants judged either aspects of the action's means or its outcome. Compared to attending to the outcome of the action, attention to means produced significantly more activation in the MNS. Similarly, Spunt et al. (2011) showed that during action observation, the goal to verbally identify rather than passively observe enhanced activation in several regions of the MNS. These findings are corroborated by the previously reviewed work from our group (Spunt & Lieberman, 2012a, 2012b), which similarly demonstrates that attention to *how* an action is performed enhances MNS activation more than attention to *why* the same action was performed. Indeed, even in the absence of observable actions, as when an action is merely described (Spunt et al., 2010), attention to *how* strongly engages MNS, demonstrating that this system can be intentionally operated. These studies strongly undermine the notion that the MNS always operates automatically. However, we do not argue that the MNS is a controlled system. Rather, in line with the dimensional–conditional approach to automaticity outlined earlier, we believe that claims to automaticity in the MNS should specify both dimensions and conditions of operation.

Discourse about the MZS often claims that this system relies on controlled processing (for a review, see Lieberman, 2010). However, as with the MNS, the operating characteristics of the MZS have received somewhat limited empirical investigation (selected studies are listed in Table 19.3). The majority of research on the MZS, insofar as it uses explicit goal manipulations to induce the representation of mental states (for reviews, see Carrington & Bailey, 2009; Van Overwalle & Baetens, 2009), can be said to establish firmly that MZS operation *is* indeed subject to intentional control. Indeed, it has been suggested that in the context of action observation, some areas of the MZS may not only be subject to control but also may themselves exert control over other regions (Spengler, Cramon, & Brass, 2009). In a recent study, we further investigated the controlled operating characteristics of the MZS. It has been suggested that the central component of controlled processing is working memory, which allows for online maintenance and serial manipulation of mental representations. Hence, one method for establishing controlled processing in the MZS is to determine whether it operates like a working memory system for social cognition. The most common method for investigating working memory is parametrically manipulating the amount of information to be held in mind. However, studies of working memory have almost exclusively investigated the maintenance and manipulation of nonsocial information, such as numbers and sensorimotor representations. In a recent study (Meyer, Spunt, Berkman, Taylor, & Lieberman, 2012), we developed a paradigm for investigating the online maintenance and manipulation of *social* representations, or *social working memory*. While undergoing fMRI, participants were asked to rank two, three, or four of their friends on a trait dimension (e.g., generosity) during a 6-second delay period. At the end of the period, they made a rank judgment (e.g., Is Rebecca the second most generous?) that could be coded as accurate or inaccurate based on ratings collected from the participant prior to the fMRI study. When they accurately ranked their friends, regions of the MZS, in particular the dmPFC, PCC/PC, and TPJ, demonstrated parametric increases in activity as a function of the number of friends to be ranked. This strongly suggests that regions of the MZS can be *intentionally* operated in a top-down fashion to manage increasing social cognitive demands. This converges with other studies showing that the dmPFC shows enhanced activation for more difficult mental state inferences (Jenkins & Mitchell, 2010) and in individuals who evidence greater degrees of mentalizing while playing a competitive game (Coricelli & Nagel, 2009).

Several studies have used either task-irrelevant stimulus manipulations or dual-task paradigms to evaluate the efficiency of MZS operation during the processing of social information. Several studies have used task-irrelevant stimulus manipulations to demonstrate that, for example, descriptions of behavior that are strongly associated with personality traits (i.e., are trait-diagnostic) elicit activation in areas of the MZS even when participants do not explicitly have the goal of making trait inferences (Mitchell, Cloutier, Banaji, & Macrae, 2006; Ma, Vandekerckhove, Overwalle, Seurinck, & Fias, 2011). Yet these studies only provide information about the *spontaneity* of MZS activation during social information processing. In order to provide information about *efficiency*, the manipulation (stimulus or goal) intended to produce MZS activation must be paired with a manipulation of the difficulty of a secondary task. To our knowledge, only two studies have successfully employed this method. In one, den Ouden, Frith, Frith, and Blakemore (2005) presented participants with written scenarios manipulated to induce mentalizing while simultaneously manipulating the presence of a secondary visual judgment task. Although the dmPFC was associated with mentalizing when participants were not under attentional load, this association was extinguished by the induction of load. In a previously discussed study from our group (Spunt & Lieberman, 2013), we used a dual-task paradigm to investigate the efficiency of the MZS during action observation. We factorially manipulated participants' comprehension goal (using the why/how paradigm

TABLE 19.3. Selected fMRI Studies Germane to the Automaticity Profile of the Mentalizing System (MZS) during Social Cognition

Study	Method	Findings	Implications
den Ouden et al. (2005)	Manipulated presence of (1) mentalizing judgments and (2) attentional load	Presence of attentional load extinguished dmPFC association with mentalizing judgments	Some MZS areas operate nonefficiently under attentional load
Mitchell et al. (2006)	Manipulated (1) trait diagnosticity of behavioral descriptions and (2) goal to form an impression or attend to the trial sequence	dmPFC activity distinguished diagnostic from nondiagnostic behaviors only when subjects had the sequencing goal	Some MZS areas spontaneously activate for trait-diagnostic behavioral descriptions
Coricelli & Nagel (2009)	Measured level of strategic mental state reasoning in a game where personal outcomes depend on others' choices	Individuals evidencing higher levels of mental state reasoning showed increased dmPFC and vmPFC activation	Some MZS areas are sensitive to amount of mentalizing across individuals
Spengler et al. (2009)	Subjects performed both a mentalizing task and a separate task demanding inhibition of automatic imitation	Mentalizing and imitation-inhibition tasks recruited common areas of mPFC and right TPJ	Some MZS areas are involved in control over automatic imitation
Jenkins & Mitchell (2010)	Subjects inferred characters' beliefs or preferences from scenarios that differed in the certainty with which such inferences could be made	Uncertain (ambiguous) mental state inferences increased dmPFC activation regardless of type of mental state	Some MZS areas track difficulty of mental state inferences
Ma et al. (2011)	Manipulated (1) trait diagnosticity of behavioral descriptions and (2) goal to passively read or actively infer trait	Trait-diagnostic descriptions activated dmPFC, left TPJ, and bilateral aTC in both the passive and active tasks	Some MZS areas spontaneously activate for trait-diagnostic behavioral descriptions
Wagner et al. (2011)	Manipulated social content in photographs while subjects made mentalizing-irrelevant judgments	Photographs of social scenes increased activation in the MZS; this effect was stronger for high trait empathizers	Spontaneous activation of MZS to social stimuli shows individual differences
Rameson et al. (2011)	Measured empathy for emotionally expressive actors viewed either under memory load or with the instruction to actively empathize	dmPFC activation was reduced under load; trait empathy was associated with mPFC activation only under load	Efficient activation of MZS under memory load shows individual differences
Meyer et al. (2012)	Manipulated mentalizing demands by having subjects rank either 2, 3, or 4 friends on a trait dimension	Increasing mentalizing demands increased activation in dmPFC, PCC/PC, and left TPJ	Some MZS areas are sensitive to increasing mentalizing demands
Spunt & Lieberman (2013)	Manipulated (1) mentalizing goal and (2) level of memory load during action observation	Increasing memory load extinguished assocation of dmPFC and left aTC with mentalizing goal	Some MZS areas operate nonefficiently under memory load

Note. aTC, anterior temporal cortex; dmPFC/vmPFC, dorsomedial/ventromedial prefrontal cortex; PC, precuneus; PCC, posterior cingulate cortex; TPJ, temporoparietal junction.

described earlier) and their level of memory load during the observation of goal-directed actions. Whereas several regions of the MNS showed a load-independent response to the actions, suggesting efficient operation, two regions of the MZS, namely, the dmPFC and the left aTC, showed a response that was strongly modulated by load only when participants had the goal to understand the actor's motives (that is, in the *why* condition). This suggests that under some conditions, the MZS can be intentionally operated upon, and that this operation is dependent on the availability of attentional resources.

Although some areas of the MZS may be critically involved in the online control of social cognition, several studies have shown that that MZS activation can be driven by stimulus variation in the absence of an explicit goal to infer mental states (Mitchell et al., 2006; Brass et al., 2007; Ma et al., 2011; Wagner, Dal Cin, Sargent, Kelley, & Heatherton, 2011; Wheatley, Milleville, & Martin, 2007), a feature that suggests efficient processing. For example, Wagner et al. (2011) had participants undergo fMRI while categorizing a series of images as containing animals, vegetables, or minerals. As a group, they observed increased MZS activation in response to photographs of humans (category: animal) compared to photographs of nonhuman animals, vegetables, or minerals. Insofar as this activation was task-irrelevant, this suggests efficient processing. Moreover, they found that individuals reporting higher levels of trait empathy exhibiting enhanced task-irrelevant activation of the MZS. As described earlier, this captures not the efficiency but the spontaneity of MZS activation. In order to capture the former dimension, the difficulty of the task-irrelevant judgment could have been manipulated to see whether the level of spontaneous MZS activation was unaffected by increasing attentional load. In a recent study, members of our group observed a similar effect in the MZS while also manipulating load. Rameson, Morelli, and Lieberman (2011) used fMRI to investigate efficiency of MZS during the perception of contextualized emotional displays. Participants viewed photographs of targets expressing emotions with one of two instructions: (1) empathize with the target or (2) watch while holding in mind an 8-digit number (an induction of cognitive load). Consistent with inefficient operation, several regions of the MZS were strongly deactivated by the induction of load. However, consistent with individual differences in efficient operation, reports of trait empathy were associated with MZS activation to the photographs when individuals were under cognitive load. High trait empathizers produced similar levels of MZS activity with and without the cognitive load. As with the MNS, these studies suggest the importance of a multidimensional and conditional view of automaticity in the MZS. Moreover, they suggest that an important condition on MZS automaticity may be individual differences in social expertise.

Summary

In this section, we have reviewed social neuroscience evidence regarding the operating characteristics of the MNS and MZS during action observation and social cognition (summarized in Tables 19.2 and 19.3). We first considered the claim that the MNS operates automatically, and concluded that while there *is* some evidence to suggest that the MNS operates *efficiently* (e.g., Iacoboni et al., 2005; Spunt & Lieberman, 2013), there is also plenty of evidence demonstrating that its operation is subject to online control (e.g., Iacoboni et al., 1999; Hesse et al., 2008; Spunt et al., 2010; Spunt & Lieberman, 2012a, 2012b). Next, we considered the claim that the MZS is a controlled processing system. Regarding this claim, there is strong evidence supporting the claim that the MZS is indeed subject to online control and may even operate in a manner that adapts to the processing demands of the task at hand (e.g., Meyer et al., 2012). Several researchers have attempted to demonstrate that the MZS operates *efficiently* in the presence of relevant social stimuli; however, a close examination of these methods shows that most claims to *efficiency* in the MZS are based on *only* the demonstration of spontaneous (i.e., task-irrelevant) activation (e.g., Mitchell et al., 2006; Brass et al., 2007; Wagner et al., 2011). This raises a point that is worth repeating: Spontaneous operation by itself does not constitute a demonstration of efficient operation. To our knowledge, the only study to test directly the efficiency claim actually showed inefficient operation of two core regions of the MZS (Spunt & Lieberman, 2013). The review further underscores the utility of a dimensional–conditional approach to neural automaticity that acknowledges multiple dimensions of operation that may be sensitive to the context (i.e., condition) of operation. Moreover, it highlights the need to scrutinize carefully the methods used to investigate the operating characteristics of a neural process.

CONCLUSION

In writing this chapter we had two basic motivations. One, we sought to apply the multidimensional framework of automaticity (Bargh, 1994) to social neuroscience research examining the operating characteristics of social information processing in the brain. We believe this provides a useful framework within which to make claims regarding the operating characteristics of neural processes. Moreover, we believe it underscores the utility of social neuroscience studies for testing hypotheses regarding the operating characteristics of the mental processes that are near and dear to social psychologists. One of these processes—the causal explanation of behavior—motivated us in the second part of this chapter, which was to critically evaluate the automaticity profiles of two brain systems known to be involved in this process. Overall, we believe our analysis emphasizes the necessity of conditional statements regarding the automaticity of a brain system. The methods of social neuroscience can shed light on the nature of automaticity and control in social cognition by allowing psychological scientists to characterize the operating characteristics of the brain systems known to support social cognition. In so doing, the goals should not be to reify a neural process as automatic but to identify the conditions under which it does (or does not) demonstrate characteristics of automatic operation. In other words, the critical question should not be *whether* the social brain operates automatically; rather, the critical question should be *when* and *for whom* it operates automatically.

NOTES

1. We recognize the the two most influential attributional process models, Trope's identification–inference model (Trope, 1986) and Gilbert's characterization–correction model (Gilbert et al., 1988) disagree regarding the automaticity profile of attributional processing. Whereas in Trope's model the attribution process is described by a single controlled process, Gilbert's model decomposes the attributional process into a sequence of two steps: automatic attribution to disposition (in line with work on spontaneous trait inference; cf. Uleman, Saribay, & Gonzalez, 2008) followed by controlled correction of the attribution based on additional information (e.g., situational information).
2. Given that single neurons exhibiting production–perception "mirroring" have not been identified in the canonical areas of the human MNS (but see Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010, for a study identifying them in other areas), some researchers prefer alternative labels such as the *mirror system*, the *motor resonance system*, or the *action observation network*.

REFERENCES

Bargh, J. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. Uleman & J. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York: Cambridge University Press.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology*, *17*(24), 2117–2121.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain?: A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–2335.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167.

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

Chong, T. T.-J., Williams, M. A., Cunnington, R., & Mattingley, J. B. (2008). Selective attention modulates inferior frontal gyrus activity during action observation. *NeuroImage*, *40*(1), 298–307.

Coricelli, G. (2005). Two-levels of mental states attribution: From automaticity to voluntariness. *Neuropsychologia*, *43*, 294–500.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519–526.

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology*, *18*(6), 454–457.

den Ouden, H. E. M., Frith, U., Frith, C., & Blakemore, S.-J. (2005). Thinking about intentions. *NeuroImage*, *28*(4), 787–796.

Desmurget, M., & Sirigu, A. (2009). A parietal–premotor network for movement intention and motor awareness. *Trends in Cognitive Sciences*, *13*(10), 411–419.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–693.

Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *Journal of Neuroscience, 30*(25), 8421–8424.

Falk, E. B., Berkman, E. T., Whalen, D., & Lieberman, M. D. (2011). Neural activity during health messaging predicts reductions in smoking above and beyond self-report. *Health Psychology*, *30*(2), 177–185.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences USA*, *103*(31), 11778–11783.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*(5722), 662–667.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1431), 517–528.

Gallese, V. (2007). Before and below "theory of mind": Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1480), 659–669.

Gallese, V. (2009). Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research*, *73*(4), 486–498.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609.

Gilbert, D. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. Uleman & J. Bargh (Eds.), *Unintended thought* (pp. 189–212). New York: Guilford Press.

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.

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience*, *24*(24), 5500–5505.

Hamilton, A. F. de C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*(4), 1133–1137.

Hamilton, A. F. de C., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, *18*(5), 1160–1168.

Hauk, O., Shtyrov, Y., & Pulvermüller, F. (2008). The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. *Journal of Physiology (Paris)*, *102*(1–3), 50–58.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Hesse, M., Sparing, R., & Fink, G. (2008). End or means—the "what" and "how" of observed intentional actions. *Journal of Cognitive Neuroscience*, *21*(4), 776–790.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), 529–535.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526–2528.

Jenkins, A., & Mitchell, J. (2010). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex*, *20*, 404–410.

Jones, E., & Harris, V. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psycholgoy* (Vol. 2, pp. 219–266). San Diego: Academic Press.

Jones, E., & Harris, V. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24.

Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Progress in Brain Research*, *156*, 379–401.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642.

Lieberman, M. (2010). Social cognitive neuroscience. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 143–193). New York: McGraw-Hill.

Lieberman, M. (2011). Why symbolic processing of affect can disrupt negative affect: Social cognitive and affective neuroscience investigations. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 188–209). New York: Oxford University Press.

Lieberman, M. D. (2007). The X-and C-systems: The neural basis of automatic and controlled social cognition. In E. Harmon-Jones & P. Winkielman (Eds.), *Fundamentals of social neuroscience* (pp. 290–315). New York: Guilford Press.

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

Luo, Q., Holroyd, T., Majestic, C., Cheng, X., Schechter, J., & Blair, R. J. (2010). Emotional automaticity is a matter of timing. *The Journal of Neuroscience*, *30*(17), 5825–5829.

Ma, N., Vandekerckhove, M., Overwalle, F. V., Seurinck, R., & Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. *Social Neuroscience*, *6*(2), 123–138.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience*, *15*(3), 394–408.

Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences*, *109*(6), 1883–1888.

Mitchell, J., Cloutier, J., Banaji, M., & Macrae, C. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience, 1*(1), 49–55.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences USA*, *96*(4), 1680–1685.

Morris, M. W., & Mason, M. F. (2009). Intentionality in intuitive versus analytic processing: insights from social cognitive neuroscience. *Psychological Inquiry*, *20*(1), 58–65.

Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience*, *31*(26), 9594–9598.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, *20*(8), 750–756.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, *14*(8), 1215–1229.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–158.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *The Journal of Neuroscience*, *23*(10), 3990–3998.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences USA*, *99*(17), 11458–11463.

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

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2011). The neural correlates of empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, *24*(1), 235–245.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Spengler, S., Cramon, D. Y. V., & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping*, *30*(11), 3704–3718.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science*, *21*(11), 1593–1598.

Spunt, R. P., & Lieberman, M. D. (2012a). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage*, *59*, 3050–3059.

Spunt, R. P., & Lieberman, M. D. (2012b). Dissociating modality-specific and supramodal neural systems for action understanding. *Journal of Neuroscience*, *32*(10), 3575–3583.

Spunt, R. P., & Lieberman, M. D. (2013). The busy social brain: Evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychological Science, 24*(1), 80–86.

Spunt, R. P., Lieberman, M. D., Cohen, J. R., & Eisenberger, N. I. (2012). The phenomenology of error processing: The dorsal ACC response to stop-signal errors tracks reports of negative affect. *Journal of Cognitive Neuroscience, 24*(8), 1753–1765.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience*, *23*(1), 63–74.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*(3), 239–257.

Uleman, J., Saribay, S., & Gonzalez, C. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Reviews of Psychology*, *59*, 329–360.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or Wrong? The Brain's Fast Response to Morally Objectionable Statements. *Psychological Science*, *20*(9), 1092–1099.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564–584.

Wagner, D. D., Dal Cin, S., Sargent, J. D., Kelley, W. M., & Heatherton, T. F. (2011). Spontaneous action representation in smokers when watching movie characters smoke. *Journal of Neuroscience*, *31*(3), 894–898.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *The Journal of Neuroscience*, *18*(1), 411–418.

Wheatley, T., Milleville, S., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, *18*(6), 469–474.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655.