available at www.sciencedirect.com

SCIENCE DIRECT®

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Mentalizing and Marr: An information processing approach to the study of social cognition

Jason P. Mitchell

Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Accepted 29 December 2005
Available online 13 February 2006

Keyword:
Social cognition

ABSTRACT

To interact successfully, individuals must not only recognize one another as intentional agents driven primarily by internal mental states, but also possess a system for making reliable and useful inferences about the nature of those beliefs, feelings, goals, and dispositions. The ability to make such mental state inferences (i.e., to mentalize or mindread) is the central accomplishment of human social cognition. The present article suggests that our understanding of how humans go about making mental state inferences will benefit from treating social cognition primarily as an information processing system that comprises a set of mechanisms for elaborating more basic social information into an understanding of another's mind. Following Marr's [Marr, D., 1982. Vision. W. H. Freeman, San Francisco, CA] framework for the study of such information processing systems, I suggest that questions about social cognition might profitably be asked at three levels – computation, algorithm, and implementation – and outline a number of ways in which a description of social cognition at the middle level (i.e., the step-by-step processes that give rise to mental state inferences) can be informed by analysis at the other two.

Understanding the behavior of other people relies heavily on the realization that humans are driven primarily by complex, internal mental states rather than by external, mechanistic forces. Since being articulated by Dennett (1987), the observation that perceivers adopt this kind of "intentional stance" in making sense of others has garnered a good deal of attention as a philosophical point. However, in order to consider others as mental agents, perceivers need more than just a good philosophy of mind; they also require a specific set of cognitive skills that convey the ability to accurately and efficiently attribute mental states to others. That is, although any hope of understanding the behavior of other people rests on first recognizing that humans are intentional agents who act primarily on the basis of their thoughts, feelings, and motivations, the actual usefulness of this recognition subsequently derives from whatever cognitive processes actually permit one human to make accurate and rapid inferences about the internal states of another.

As such, the primary goal of research on social cognition is the illumination of the nature of the cognitive mechanisms that support inferences about the mental states of others. However, this simple statement about the nature of social cognition and the goals of those who study it obscures several difficulties in unpacking exactly what researchers of social cognition are (or should be) doing. Two major questions are intrinsic to this enterprise. First, one can ask whether researchers need to start by postulating the existence of special mechanisms devoted to social cognition or whether an understanding of others' mental states can be accomplished on the basis of more general cognitive processes. This question is neatly summed up by Blakemore et al. (2004), who recently asked whether "the general cognitive processes

E-mail address: mitchell@wjh.harvard.edu.

0006-8993/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2005.12.113

involved in perception, language, memory and attention, are sufficient to explain social competence or, over and above these general processes, are there specific processes that are special to social interaction?" (p. 216).

Second, what kind of "illumination" of social cognition should researchers be seeking? What does it even mean to "understand" a cognitive system? Do researchers want a neuron-by-neuron account of how social cognition is carried out by the human brain?; the blueprint for programming a computer to mentalize successfully?; some other higher-order theory of the nature of social cognition? As researchers of social cognition, we need to know where we are headed scientifically and how to recognize our destination should we happen to arrive there.

1. Social cognition and Marr's information processing framework

In this article, I would like to suggest that the answers to these questions can be found in a consideration of social cognition, first and foremost, as an *information processing* system. An information processing system is one that, generally speaking, takes raw input and subjects it to a series of transformations that render the output more complex or more useful than the original information. For example, computers are classic information processing systems, taking certain kinds of input (e.g., keyboard presses) and transforming them into more elaborate, useful behaviors (a cut-and-paste command; a network transmission of an electronic message; etc.). Since the overthrow of behaviorism by a cognitivist approach to psychology, the human mind has likewise been viewed as a complex information processing system, capable of transforming fairly simple information (e.g., two-dimensional retinal input) into more useful, elaborated output (a conscious visual experience of the world around us).

An approach that treats social cognition as an information processing system entails viewing it as a means for representing basic information about the social environment and systematically transforming that information in ways that make explicit what we want to know about the other minds around us. Such an approach rests on the assumption that our inferences about others' minds do not arise ex nihilo but instead must be actively constructed out of the more basic information that we glean about others. The overarching goal of research in social cognition is therefore exposing what that basic information is as well as the transformations to which it is subjected.

More than two decades ago, David Marr (1982) outlined a powerful framework for organizing precisely these kinds of questions about information processing systems. Surveying the way in which researchers had gained traction on the specific problem of seeing the physical world – i.e., vision – Marr suggested that cognitive scientists could address their questions about information processing systems (such as the human mind) at three distinct hierarchical levels of analysis, which he described as those of *computation*, *algorithm*, and *implementation*. The highest level of analysis – that of the computation – asks about *what* problem the information processing system has been designed (or has evolved) to tackle and *why* a particular strategy represents a viable solution to the problem. To illustrate this level, Marr offers the example of a cash register, which has a clearly defined computational goal of taking numerical inputs and subjecting them to a particular arithmetic transformation (i.e., addition) to create a novel re-representation of the original information (i.e., a sum). The fact that a cash register adopts the particular arithmetic operation of addition follows directly from the fact that addition turns out to be the best strategy for achieving the computational goals of the cash register (in fact, there is such a tight link between the computational goal of the cash register and the theorem of addition that, for all practical purposes, the two are identical).

If the computational level of analysis asks "what" an information processing system does and "why", Marr's second level – that of the algorithm – asks "how" the system goes about getting the job done. Even after one knows what a system "is for," a question still remains regarding the step-by-step recipe for accomplishing those tasks. For example, although the cash register likely accepts inputs in a decimal number system, computer arithmetic is more easily accomplished using a binary system. Does the cash register go through the extra effort of re-representing numbers in binary or does it simply operate over the decimal values? Either algorithm will allow the register to meet the computational goal of adding numbers, so a full understanding of the machine requires specifying which particular information processing algorithm it happens to use.

Lastly, Marr points out that such algorithms must actually be executed by a physical system and that, as such, questions regarding a system's implementation make up a necessary third level of analysis. The calculations performed by a cash register can be implemented on a microchip or on an abacus, and the goal of this third kind of analysis is to understand the "hardware" being used to run the particular information processing algorithm. Because different hardware can more or less readily support various processing algorithms (the number of digits on the human hand makes it relatively easy for us to use base-10 or base-5 number systems), understanding the hardware used for information processing can provide important constraints on the architecture of the algorithms likely to be in use (Kosslyn and Maljkovic, 1990).

So, what does all this have to do with the study of social cognition? To the extent that the fundamental problem faced when thinking about other people is primarily one of generating elaborated inferences about another's mental states on the basis of less complex social information (such as physical cues that a person may be conveying), social cognition can be approached as an information processing problem subject to analysis at Marr's three levels of computation, algorithm, and implementation. What is the overarching computational goal of social cognition? That is, what strategies are useful for mentalizing about others? Second, by what algorithmic process are such strategies realized? Finally, what underlying brain regions are used to implement these processes? As for cognitive research into other domains (such as vision), a satisfactory psychological understanding of social cognition requires answers to all three types of questions.

Although Marr (1982) argued that the three levels were independent and could therefore be studied without reference

to each other, subsequent commentators (Kosslyn and Maljkovic, 1990) have traced how Marr's cognitive theories about vision (at the level of the algorithm) were informed heavily by analysis at the other two levels. In the remainder of this article, I will focus on how analysis of social cognition at two of the levels – those of the computation and implementation – can help guide both the questions asked as well as the answers provided at the remaining level, that of the cognitive algorithm. The goal here is therefore two-fold: first, to suggest how consideration of the computational problems faced by mental state inference can constrain the questions researchers ask about the underlying algorithms of social cognition and, second, to suggest ways in which illuminating the neural basis of social cognition has already begun to provide an emerging sketch of the mechanisms subserving human social abilities.

2. A computational analysis of social cognition

2.1. *Social cognition as unitary or manifold?*

In Marr's view, the highest-level question to ask about an information processing system is what that system is *for*. What is the purpose of the system, what computational problem has it evolved (or is it designed) to solve? Even an initial answer to these questions first requires correctly identifying the computational level at which the system contributes to information processing. For example, although the purpose of vision is clearly to provide a representation of the physical environment surrounding an organism, many other systems also contribute their own, independent representations of the physical world: audition, proprioception, electroreception (at least in some fish), etc. Vision can therefore be considered one solution – from a suite of solutions – to the problem of representing the physical world. For a computational analysis of these problems to be achieved, one must first decide how best to frame the question of interest: does one want to inquire about how a representation of the physical world is achieved in general (which would likely entail a consideration of how information from different sensory systems is integrated into a coherent representation) or specifically about how one part of that overall representation is provided by the visual system?

Similarly, the questions associated with social cognition first need to be given a specific framing. In much the same way that perception provides a representation of the relevant properties of our physical environment, social cognition can be thought of as providing a representation of the relevant properties of our social environment, a large part of which will comprise the mental states of other people. As such, the computational question of social cognition in general is analogous to the computational question of perception in general (not to one particular perceptual system in particular) and subdivides similarly into two parts. First, what are the individual parts of the overall system for mental state inference, and what does each of these subcomponents contribute to an overall representation of the social environment? Second, how is the information provided by these subcomponential systems integrated to provide a single, coherent representation of the social world? Not only are these two questions separable, but demarcating this separation is necessary for knowing how to address the computational analysis of social cognition in the first place.

The differences between two kinds of computational questions neatly separate two contemporary approaches to study of social cognition, one of which treats representing the social word as the function of a single, unitary system (akin to vision), whereas the other considers "social cognition" an umbrella term referring to a set of related processes that divy up various aspects of handling the social world (more like the broad idea of "perception"). On this first view, the prevailing theoretical approach to the problem of inferring others' mental states has been to frame questions around a single system for representing the social world, without giving much consideration to the way in which this system could be subdivided into separate cognitive processes. This approach has been embodied in arguments regarding a "module" dedicated to mentalizing, most closely associated with Alan Leslie (Leslie, 1992, 1994; Leslie et al., 2004), who has argued that our mentalizing abilities rely on the operation of a unitary theory of mind mechanism (abbreviated as ToMM). Although this hypothesized module would necessarily receive input from lower-order processes common to many tasks (such as visual input about where someone is looking), ToMM is thought to bear the brunt of the work necessary to understand others' mental states in a unitary, encapsulated way. Similar logic guides the array of other work that has attempted to identify a single mechanism underlying mental state inference (Carruthers, 1996; Gopnik and Wellman, 1992, 1994; Saxe, 2005). By suggesting that social cognition is the purview of a single cognitive process, these views obviate the need to consider any additional, higher-order mechanisms of integrating distinct aspects of social thought.

On the other hand, recent commentators have begun to take seriously the view that social cognition cannot be treated as a single, specialized module but instead needs to be thought of as a consortium of specialized cognitive processes, each of which solves some circumscribed aspect of the overall problem of mentalizing (Ames, 2005; Macrae et al., 1994; Malle, 2005). These theorists have developed the metaphor of social cognition as a "toolbox," wherein a set of different cognitive processes might all be utilized for mental state inference, depending on the particulars of the social situation one encounters. For example, one-on-one interactions provide cues about another's mental state that are unavailable in telephone conversations, such as eye gaze, emotional expression, hand gestures. On the other hand, the telephone affords cues – such as prosody and other pragmatic aspects of spoken speech – not available in situations where one person attempts to communicate an intention silently to another (as experienced by anyone who has ever wanted to communicate a desire to leave a dinner party to a spouse without offending the host). Likewise, some researchers have recently argued that one potentially useful approach to understanding another's mind is by reference to one's own (sometimes referred to as simulation or projection), but only for people whom one can assume are relevantly similar to oneself (Ames, 2004a,b; Mitchell et al.,

2005b). Accordingly, distinct cognitive mechanisms may be in place for mentalizing about similar and dissimilar others since the social–cognitive "tool" of simulation/projection is appropriate to one group but not the other. Successful mentalizing demands the flexible ability to use each of these cues when available without relying on the presence of any one. Unlike more unitary views of social cognition, this "toolbox" approach will also require understanding the processes by which distinct social knowledge (e.g., a target's facial expressions, tone of voice, and past behavior) is integrated into a coherent representation of another's mind, although little is currently known about the nature of such integrative processes.

2.2. *The constituent parts of social cognition*

Support for this latter view of social cognition as "manifold" – i.e., divisible into component parts that each tackles a specific aspect of mentalizing – derives from any computational analysis of social cognition that starts from the question of what information could possibly be used to uncover the hidden mental states of others. In considering the case of human vision, Marr (1982) pointed out that the computational goal of the visual system has been necessarily constrained by what information presents itself in the actual, real-world perceptual environment (also see Gibson, 1979). Although many strategies might have given rise to the richness of human visual experience, the human visual system appears to rely heavily on a specific computational strategy of extracting information about object surfaces. Presumably, the strategy of "seeing" through recovery of surface information is used by humans because it satisfies the twin requirements of being both *useful* (it provides reliable information about the actual state of the physical world that can be acted on) as well as *useable* (in actual practice, the visual system has information about surfaces available to it since these are the aspects of objects that make themselves visible to the retina).

So what features of the social environment are available for social cognition to exploit for the purposes of mental state inference? More specifically, what reliable useful cues do targets convey about their mental states? A partial list would no doubt include eye gaze (where or at whom another person is looking); facial expression (a slightly wry smile, raised eyebrow, or full-blown expressions of emotion); body posture and gait (e.g., compare a slouching, "defeated" walk to a confident stride); gesturing and other micro-movements (head nodding, yawning, leg shaking); paralinguistic aspects of speech, such as tone of voice and fluency of speech; and the actual content of speech itself (hearing someone say "I'm hungry" is a powerful cue to his or her internal state).

In addition, it is also likely that perceivers use some kind of top–down processing to amplify the meaning of such cues or to provide additional information not contained within the input provided by a target. Perceivers may make use of precomputed knowledge about the link between certain situations and the mental states provoked by them, which may, under certain conditions, entirely support a mental state inference (e.g., people feel pain when they are kicked in the shins). In addition, our social knowledge tends to include beliefs that certain mental states are more likely for certain groups of people than others (that cheerleader will feel more pain than that soccer player after a shin-kicking), and research by Leyens and colleagues (Demoulin et al., 2004; Leyens et al., 2000; Paladino et al., 2002) has demonstrated that perceivers are less willing to acknowledge that members of outgroups have the kind of rich mental experiences that ingroup members do. Finally, as mentioned above, one can potentially use knowledge about one's own mental states to infer those of others: to the extent that a perceiver believes that another person thinks, feels, and wants the same things in the same situations as herself, she can use predictions about what she would think, feel, or want to support inferences about that other person's thoughts, emotions, and desires (Adolphs, 2002; Davies and Stone, 1995a,b; Gallese and Goldman, 1998; Gordon, 1992; Heal, 1986; Meltzoff and Brooks, 2001; Mitchell et al., 2005b; Nickerson, 1999).

Such an ecological approach assumes that social cognition will primarily comprise a set of opportunistic mechanisms for capitalizing on the kinds of useful social information that presents itself in the actual social environment (McArthur and Baron, 1983). This view necessarily raises a host of questions regarding the various processes by which different kinds of mental state inferences may be made. First, is it really the case that each of these avenues for mental state inference is serviced by a distinct set of cognitive processes? What little is known regarding this question has sometimes suggested fairly counterintuitive answers: whereas different aspects of the same speech stream – e.g., semantic content vs. prosody – are processed very differently (indeed, by different hemispheres of the brain), there may be considerable overlap in the mechanisms used to understand eye gaze, emotional expression, and gesture (for a review, see Allison et al., 2000). Deeper analysis of the kinds of information that need to be recovered from each of these social–cognitive cues would help provide a more distinct outline of the nature (and number!) of the algorithms subserving each.

Second, one can ask about the particular situations in which each of these hypothesized tools for mentalizing might be put to use. As mentioned above, many cues are provided in some contexts but not others, and any social–cognitive system worth its salt must be able to make flexible use of whatever information presents itself in the social environment. Over and above these constraints, however, there are distinct differences between making online mental states inferences (what is this person feeling/thinking right now?) and making predictions about someone's potential future mental states. Whereas online mental state inferences may be supported by any number of immediately available cues being transmitted by a target, predictions about the future (e.g., how would my friend feel if I did not show up at her party?) must generally be accomplished in the absence of such physical cues and may therefore rely more heavily on the "top–down" social knowledge that perceivers maintain about the way other people work. For the most part, few attempts have been made to separate such online judgments from more predictive aspects of social cognition, although a full account of social cognition requires a rich understanding of the different processes used to make each of these kinds of mental state inferences.

2.3. *What good is social cognition?*

One final aspect of a computational analysis of social cognition asks to what use mentalizing is put. In Marr's (1982) consideration of vision, the answer to this "what purpose" question is inextricably linked to the behavior of the organism in question. The visual system of the fly comprises just a handful of simple mechanisms presumably because the demands on a fly's visual system are sufficiently small to allow getting by with such a bare-bones system. The visual system of the frog is considerably more complicated because frog behavior is considerably more complex than fly behavior, but still much less so than that of primates (who have a concomitantly more complicated visual system). Marr's takeaway message would seem to be that the nature of the particular kind of perceptual representation used by an organism is determined by what that organism generally does with such representations.

In much the same way, social cognition must provide perceivers with a representation of other minds that is useful for guiding the perceiver's own behavior. Ultimately, knowing that other humans are intentional agents and even understanding the content of their mental states are not enough to make social cognition useful without the additional capacity to guide what a perceiver actually does. Among the most important aspects of this requirement is to use mental state inferences (or predictions) to figure out how to behave in a way that successfully influences another's mental states. Our social–cognitive abilities are frequently deployed for the purpose of figuring out how to manipulate what another person will think or do in response to what we ourselves think, do, or say (Byrne and Whiten, 1988). Indeed, almost any speech act is inherently an attempt to manipulate another person's mental states. For example, this article can be viewed as my own attempt to induce particular mental states in the reader – beliefs about the nature of social cognition – and doing so demands a complex prediction about how a reader's thoughts will change in response to my own behavior (in this case, writing).

This whirlwind overview of the computational issues posed by social cognition obviously raises more questions than it addresses. Nevertheless, as for other domains of psychological inquiry, even such brief and cursory outline can provide a kind of roadmap for the kind of questions to ask about a domain of psychological inquiry. For social cognition, those questions include consideration of whether human mindreading abilities rely on cognitive processes that distinguish it from other forms of thought; whether social cognition is best approached as a unitary module for solving the problem of mental state inference or a collection of related subprocesses; what kinds of information perceivers bring to bear to infer another's mental states; and a consideration of the interplay between inferring and manipulating others' mental states.

3. The implementation of social cognition

If Marr's first level of analysis can help frame the questions that researchers ask about the nature of social cognition, analysis at the third level – that of implementation – promises to provide empirical constraints on the answers at which researchers arrive. Although there is no doubt that human information processing is primarily implemented on a single piece of biological hardware – the brain – using our understanding of the "hardware" of social thought to constrain or inspire theories at the level of the cognitive algorithm remains a sizeable challenge. However, Marr's own work on the cognitive algorithms underlying visual perception was heavily informed by analysis of how vision was implemented, in particular, with regard to the receptive field properties of neurons in different regions of the brain (Kosslyn and Maljkovic, 1990). Can a consideration of the brain basis of social cognition likewise help inform cognitive theories about how we think about the minds of others?

Recently, Henson (2005) has outlined the two assumptions necessary to use data about where in the brain psychological processes are instantiated to, in turn, constrain theories about the nature of the those processes themselves. First, different brain regions must be assumed to subserve different cognitive processes and, second, a single brain region must subserve only a single such cognitive process. These two assumptions allow brain localization research to contribute to our understanding of information processing at the level of the cognitive algorithm in two ways. First, one can interpret an observation that two tasks yield different patterns of brain activity as evidence that the two tasks provoke different kinds of cognitive processing. Inversely, one can interpret an observation that two tasks engage overlapping patterns of brain activity as evidence that the two tasks share whatever cognitive processes are supported by those brain regions. This section reviews extant research that has adopted these assumptions about the relation between brain and mind (i.e., between implementation and algorithm) to illuminate the cognitive basis of social cognition.

3.1. *The "distinctiveness" of social cognition*

Recent analysis of social cognition at the level of the implementation has made use of both assumptions to augment current understanding of social processing. The first of these provides a direct answer to the question raised by Blakemore et al. (2004) regarding whether inferring the mental states of other people draws on a distinct set of cognitive mechanisms or simply "piggybacks" on the same general-purpose processes important for other kinds of information processing (e.g., memory, attention, semantics, inferential reasoning, etc.). In recent years, an increasing number of brain localization studies have been weighing in on the side of the former possibility. Using both neuroimaging and neuropsychological patients, researchers have consistently observed a set of neuroanatomical regions – medial prefrontal cortex (mPFC), superior temporal sulcus (STS), and lateral parietal cortex – that distinguishes social cognition from other complex cognitive tasks. By showing differences in the brain basis of social and nonsocial cognition, these studies have capitalized on Henson's first assumption (i.e., different brain regions imply different cognitive processes) to suggest that thinking about the minds of other people relies on a distinct

set of cognitive algorithms that separates social cognition from other forms of thought.

Some of the first brain data to suggest the distinctiveness of social cognition came from work by Goel et al. (1995), who compared patterns of neural activation observed when participants were asked to indicate whether a historical figure (Christopher Columbus) would correctly recognize the function of various artifacts (such as a compact disc) to neural activation when asked to consider semantic or visual aspects of those objects. Results indicated that inferring what another person knows led to greater activation in a set of regions that included mPFC and STS. Similar observations were made around the same time by Fletcher et al. (1995), who observed greater mPFC and STS activation when participants read stories that could be understood only through a consideration of the mental states of the characters compared to stories that simply required an understanding of physical causality. A follow-up study (Gallagher et al., 2000) replicated and extended these results to include mental state attribution made for nonverbal stimuli (cartoons).

More recent work has successively strengthened the basis for the belief that these patterns of brain activation specifically result from the requirement to make inferences about mental states. For example, a number of studies have asked participants to play interactive games that require second-guessing one's opponent, such as in the children's game "rock, paper, scissors" (Gallagher et al., 2002) or a decision-making "trust" game (McCabe et al., 2001). These researchers compared neural activity when subjects thought they were playing against a human opponent to activations when subjects thought they were playing against a computer. In each case, although the tasks were otherwise identical, situations in which participants relied on mental state attribution to compete against their opponent (because they believed they were playing against another human instead of a computer) led to activation of the mPFC, lateral parietal cortex, and superior temporal sulcus.

In a series of recent experiments, we have observed similar dissociations even when the identity of targets is not manipulated (participants always judge other people), but the requirement to mentalize varied. For example, mPFC was differentially engaged by (1) judging the mental state of a photographed person compared to judging the symmetry of that person's face (Mitchell et al., 2005b); (2) using experimentally provided information to form an impression of another person compared to intentionally encoding the order in which that information was presented (Mitchell et al., 2004, 2005c); and (3) judging the potential mental characteristics of people compared to judging their potential physical aspects (Mitchell et al., 2002, 2005a).

A similar dissociation was reported by Kumaran and Maguire (2005), who compared mental navigation of the physical world to mental "navigation" of the social world. Participants in this study were asked to report how they would transfer a crate of wine from one friend to another in London using either a physical route between the locations of the friends' residences or by asking one person to pass it along to another in the participant's extended social network, which would require reference to whether the targets actually know each other. Consistent with earlier work, considering the social relationships among one's friends engaged regions of the mPFC, lateral parietal cortex, and superior temporal sulcus over and above considering their relative physical relationships. Together, these data demonstrate that the social–cognitive processes subserved by these regions are not blindly engaged in the presence of another person but are specific to situations in which the mental states of others must be inferred.

Finally, Saxe and her colleagues (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005) have reported a set of studies comparing patterns of brain activation when perceivers consider the beliefs of other people to activation when perceivers consider equivalently complex nonsocial aspects of a situation. For example, in one study, participants alternately considered "false belief" problems in which a protagonist's beliefs conflicted with the actual state of the world (for example, someone might believe that a car was a Porsche when it was really a Ford) or "false photograph" problems in which a photograph depicted a scene that conflicted with the actual state of the world (because something had happened to change the scene since it was photographed). Despite the formal similarity between these two tasks (both require inferences about a mismatch between reality and a representation of it), differential engagement of mPFC and lateral parietal cortex (at the junction of temporal and parietal cortices) was associated with the mentalizing task, again suggesting that distinct processes are deployed during social and nonsocial reasoning.

Together, these brain localization data suggest various ways that analysis at the level of the implementation can weigh in on questions about the algorithms of social cognition. By demonstrating the distinctiveness of mental state inference from other forms of processing, these data provide a particular answer to the question asked by Blakemore et al. (2004), namely, that mental state inferences draw on cognitive processes that are distinct from those subserving nonsocial aspects of thought. Of course, this is not to be taken as saying that social cognition makes exclusive use of its own distinct set of processes; at the very least, social cognition must rely on some of the same basic perceptual processes as other cognitive abilities (for example, it would be surprising if primary visual cortex were not important both for seeing other people's faces as well as inanimate objects), and it remains an open question as to how much higher-order cognitive processes contribute to social cognition (e.g., are the same mechanisms of executive control engaged in social as nonsocial situations?). At the same time, however, these data do make the more tightly circumscribed points that predicting the behavior of other people does not appear to rely on the same set of cognitive processes as predicting the behavior of inanimate objects (such as computers); that considering the mental aspects of another person does not rely on the same processes as considering his physical characteristics; that piecing together a coherent story about another's mental states does not rely on the same processes as piecing together a story about the physical state of the world; and that considering the social relation of one person to another does not rely on the same processes as considering the physical relation between those two people. In short, although social cognition will no doubt draw on a wide set of perceptual and

cognitive processes, in those studies in which head-to-head comparisons have been made between mental state inference and other complex cognitive tasks, social cognition does not seem merely to piggyback on existing cognitive processes. Instead, as suggested by functional neuroanatomical dissociations from comparable nonsocial tasks, social cognition appears to engage in an entirely different set of cognitive algorithms.

3.2. *How is mentalizing accomplished?*

Although the data discussed above certainly suggest the distinctiveness of social cognition from other forms of thought, it remains to be determined exactly what these social–cognitive processes are. In addressing this question, recent analysis of social cognition at the level of the implementation has successfully begun to outline the contours of the cognitive processes that give rise to our mentalizing abilities. These studies have capitalized on the twin strengths of neuroimaging research, demonstrating dissociations between different aspects of social cognition as well as overlap between social cognition and other kinds of cognitive processing. Along the first line, Saxe and colleagues (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005) have used fMRI to suggest that the cognitive processes deployed during mentalizing depend on what aspects of another person's mind need to be inferred. Specifically, these researchers have demonstrated that a functionally defined region at the junction of the temporal and parietal cortices (TPJ) appears to contribute specifically to making inferences about what another person *believes* (both false and true beliefs), but not other aspects of his or her mental states (such as feelings or visceral sensations, such as hunger). These results dovetail with developmental research, suggesting that one's understanding of beliefs follows a developmental timecourse distinct from understanding other kinds of mental states (Saxe et al., 2004).

Along the second line, our group has used neuroimaging to suggest an overlap between the processes used to judge one's own mental states and those used to judge those of another person. As discussed above, simulation/projection views of social cognition suggest that, when required to infer another's mental states in a particular situation, a perceiver can (consciously or unconsciously) imagine what she herself would believe, feel, or desire in the same situation and then, accordingly, assume that the other person will believe, feel, or desire roughly the same thing. However, the success of such a simulation/projection strategy obviously relies on a perceiver's confidence that she thinks and feels in a way similar to the other person in question. As such, any system that contributes to mental state inference by making reference to oneself should be sensitive to the perceived similarity between self and the target of such inferences. In two recent studies (Mitchell et al., 2005b, submitted for publication), we have investigated the response properties of a region of ventral mPFC that previous studies have linked to tasks that require self-referential thought (such as judging how well a series of adjectives or statements describe oneself). Activity in this region was also increased by mentalizing about another person but, importantly, only for others that were perceived as being similar to oneself. These brain data help suggest one mechanism of social–cognitive processing: use of self-reflection to guide inferences about others' mental states. Moreover, in both studies, a more dorsal region of mPFC was maximally engaged by mentalizing about dissimilar others, suggesting that the particular processes used for social cognition may vary as a function of the particular targets whose minds need to be figured out. This kind of brain localization approach has begun the work both of parcellating social cognition into its various processing subcomponents as well as of identifying the kinds of processing of which social cognition may consist.

3.3. *Social cognition by default*

Finally, the unique benefits gained by researchers examining social cognition at the level of implementation are perhaps most clearly revealed by the link between social–cognitive processing and the "default" state of human brain activity. Raichle and his colleagues (Gusnard et al., 2001; Gusnard and Raichle, 2001; Raichle et al., 2001; Shulman et al., 1997) have reported a curious observation that some brain regions maintain higher baseline activity than average (as measured by resting metabolic rates). Interestingly, the bulk of these regions are ones that have been associated with social–cognitive tasks. The possible implication of this observation is that social cognition seems to form part of a default mode of brain activity in which thinking about the minds of other people may be a potentiated or chronically active type of human information processing.

Because the implication of these observations is more often misunderstood than not, it is worth briefly outlining the logic underlying the arguments about a default state of human cognition (for a fuller account, see Gusnard and Raichle, 2001). The first step of the argument is a reminder that the blood-oxygenation-level-dependent (BOLD) signal used in functional magnetic resonance imaging (fMRI) reflects changes in the ratio of glucose to oxygen in the local bloody supply. Specifically, the BOLD signal arises because transient increases in neuronal activity lead to a concomitant increase in local blood supply but, critically, glucose is removed more quickly from the blood than oxygen (again, for a fuller account of the underlying biological mechanisms, see Gusnard and Raichle, 2001). The resulting decrease in the ratio of glucose to oxygen is registered as an "activation" by the fMRI BOLD signal.

Somewhat surprisingly, when individuals are permitted to rest quietly during scanning, the ratio of glucose to oxygen returns to the same equilibrium throughout the brain. Most critically, although the overall amounts of glucose and oxygen differ across brain regions, their ratio is roughly the same throughout the brain. This last point is central because it suggests that, whatever cognitive processes may be taking place when the brain is at rest, the system is set up to perform them without additional perturbation. That is, because no decoupling of glucose to oxygen takes place during rest, no fMRI "activations" are produced in conjunction with the ongoing stream of cognitive processing taking place during rest. Instead, the system appears to have evolved in such a way that certain brain regions have naturally high rates of metabolic activity, which presumably subserves a set of ongoing cognitive processing that takes place as part of the

backdrop of neural activity (perhaps somewhat analogously to the way that maintaining tonus is part of the baseline metabolic activity of skeletal muscle, separate from transient requirements to produce movements).

Gusnard and Raichle (2001) have identified four brain regions that have particularly high resting metabolic rates: dorsal mPFC, ventral mPFC, lateral parietal cortex, and medial parietal cortex (precuneus). As reviewed above, lateral parietal cortex (which includes TPJ) and both dorsal and ventral aspects of the mPFC have been consistently associated with social–cognitive processing used for making mental state inferences (the function of the fourth region, in medial parietal cortex, is poorly understood, although some commentators have suggested a role for this region in self-referential thought (Kircher et al., 2000, 2002; Newen and Vogeley, 2003; Vogeley and Fink, 2003). Although the functional significance of the high resting activity of the mPFC and lateral parietal cortex has not been worked out in its entirety, it appears likely that social cognition makes up a sizeable component of the default state of human cognition. The regions with the highest resting metabolic rates in the brain are almost uniformly those that have been associated with the social–cognitive demands of mental state inference. One view of these observations suggests the centrality of social cognition to human mental experience: when left to its own devices, the human brain appears naturally to engage in social–cognitive thought, perhaps indicating a readiness to perceive entities in the world as other mental agents.

That the human mind naturally engages in social–cognitive processing is further supported by a second curious feature of the brain regions associated with social cognition. When participants are asked to perform tasks that do not include a social–cognitive component, activity in these high-metabolism regions typically decreases; indeed, one of the most reliable observations in neuroimaging research is that activity in mPFC and lateral/medial parietal cortex is attenuated during nonsocial tasks relative to a low-level fixation baseline (producing the fMRI "deactivations" commonly observed in these areas). But why should activity be attenuated in regions subserving social cognition during nonsocial tasks? After all, the reverse effect is not commonly observed: activity associated with "nonsocial" cognitive processing does not decrease during tasks that do not require such processing (for example, nonverbal tasks do not produce deactivations in language-related areas). That regions associated with social cognition uniquely deactivate during nonsocial tasks suggests that the ongoing activity in these regions is in some way incompatible with the performance of nonsocial tasks and must be actively inhibited or suppressed. One speculative interpretation of these data suggests that perceivers naturally adopt an intentional stance in interacting with the world but that this tendency to approach entities as mental agents must be suspended in order to interact appropriately with inanimate objects that do not actually experience mental states. For example, to use a coffee mug appropriately, one has to be willing to treat it as an entity without mental states (for example, it will not feel pain when in contact with boiling liquid). If humans do indeed have an overriding inclination to approach the world as if it were full of mental agents (as the high resting metabolic rate of "social–cognitive" regions suggests), then this social–cognitive processing default will need to be actively suspended during engagement with nonmentalistic objects (as suggested by the frequent deactivations observed in these regions during nonsocial tasks).

4. Conclusion

Despite its centrality to everyday life, social cognition remains one of the most poorly understood cognitive systems. Exactly how one person ever manages to gain access to another person's mind through mental state inference remains a deep and fascinating question in psychological science. Here, I have suggested that considering social cognition an information processing system – subject to the analysis strategies first laid out by David Marr (1982) – can help highlight the approaches that will prove most fruitful for investigating issues of this kind. Our understanding of social cognition at Marr's middle level of analysis – the algorithm by which perceivers transform basic social information about another person into a representation of his or her mind – can be guided by analysis at the two adjacent levels. In particular, research on social cognition will benefit from the kind of "roadmap" provided by consideration of the set of computational problems that our social–cognitive system has evolved to solve. Moreover, the recent surge of interest in understanding the neural basis of social cognition has begun to demonstrate implementational constraints on possible social–cognitive algorithms, for example, by suggesting the distinctiveness of social cognition from other forms of thought and the parcellation of social cognition into a number of functionally discrete subcomponents (instead of a single, unitary system that subserves "theory of mind"). Investigations at the level of implementation have also suggested the surprising possibility that social cognition constitutes a large part of the default state of human neural activity, giving rise to the speculation that the algorithm of social cognition may enjoy special priority in the mind's processing stream. Future research into these issues at all three of Marr's levels of analysis promises to continue revealing precisely how the astonishing magic act of reading another's mind is accomplished.

Acknowledgments

The author is grateful for the many helpful comments provided by Hedy Kober, Kevin Ochsner, Yaacov Trope, and an anonymous reviewer.

REFERENCES

Adolphs, R., 2002. Neural systems for recognizing emotion. Curr. Opin. Neurobiol. 12 (2), 169–177.

Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cogn. Sci. 7, 267–278.

Ames, D.R., 2004a. Inside the mind reader's tool kit: projection and stereotyping in mental state inference. J. Pers. Soc. Psychol. 87 (3), 340–353.

Ames, D.R., 2004b. Strategies for social inference: a similarity contingency model of projection and stereotyping in attribute prevalence estimates. J. Pers. Soc. Psychol. 87 (5), 573–585.
Ames, D.R., 2005. Everyday solutions to the problem of other minds: which tools are used when? In: Malle, B.F., Hodges, S.D. (Eds.), Other Minds: How Humans Bridge The Divide Between Self and Other. Guilford Press, New York, pp. 158–173.
Blakemore, S.J., Winston, J., Frith, U., 2004. Social cognitive neuroscience: where are we heading? Trends Cogn. Sci. 8 (5), 216–222.
Byrne, R.E., Whiten, A., 1988. Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Oxford Univ. Press, Oxford, UK.
Carruthers, P., 1996. Simulation and self-knowledge: a defence of theory-theory. In: Carruthers, P., Smith, P.K. (Eds.), Theories of Theories of Mind. Cambridge Univ. Press, Cambridge, UK, pp. 22–38.
Davies, M., Stone, T. (Eds.), 1995a. Folk Psychology: The Theory of Mind Debate. Blackwell Publishers, Oxford, UK.
Davies, M., Stone, T. (Eds.), 1995b. Mental Simulation: Evaluations and Applications. Blackwell Publishers, Oxford, UK.
Demoulin, S., Torres, R.R., Perez, A.R., Vaes, J., Paladino, M.P., Gaunt, R., et al., 2004. Emotional prejudice can lead to infra-humanisation. In: Stroebe, W., Hewstone, M. (Eds.), European Review of Social Psychology, vol. 15. Psychology Press/Taylor and Francis, Hove, England, pp. 259–296.
Dennett, D.C., 1987. The Intentional Stance. MIT Press, Cambridge, MA.
Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., et al., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57 (2), 109–128.
Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 38, 11–21.
Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. NeuroImage 16 (3 Pt. 1), 814–821.
Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2 (12), 493–501.
Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston.
Goel, V., Grafman, J., Sadato, N., Hallett, M., 1995. Modeling other minds. NeuroReport 6 (13), 1741–1746.
Gopnik, A., Wellman, H.M., 1992. Why the child's theory of mind really is a theory. Mind Lang. 7 (1), 145–171.
Gopnik, A., Wellman, H.M., 1994. The theory theory. In: Hirschfeld, L.A., Gelman, S.A. (Eds.), Mapping the Mind: Domain Specificity in Cognition and Culture. Cambridge Univ. Press, New York, NY, pp. 257–293.
Gordon, R.M., 1992. Folk psychology as simulation. Mind Lang. 1, 158–171.
Gusnard, D.A., Raichle, M.E., 2001. Searching for a baseline: functional imaging and the resting human brain. Nat. Rev., Neurosci. 2, 685–694.
Gusnard, D.A., Akbudak, E., Shulman, G.L., Raichle, M.E., 2001. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 4259–4264.
Heal, J., 1986. Replication and functionalism. In: Butterfield, J. (Ed.), Language, Mind and Logic. Cambridge Univ. Press, Cambridge, UK.
Henson, R., 2005. What can functional neuroimaging tell the experimental psychologist? Q. J. Exp. Psychol., A 58 (2), 193–233.
Kircher, T.T., Senior, C., Phillips, M.L., Benson, P.J., Bullmore, E.T., Brammer, M., et al., 2000. Towards a functional neuroanatomy of self processing: effects of faces and words. Brain Res. Cogn. Brain Res. 10 (1–2), 133–144.
Kircher, T.T., Brammer, M., Bullmore, E., Simmons, A., Bartels, M., David, A.S., 2002. The neural correlates of intentional and incidental self processing. Neuropsychologia 40 (6), 683–692.
Kosslyn, S.M., Maljkovic, V., 1990. Marr's metatheory revisited. Concepts Neurosci. 1 (2), 239–251.
Kumaran, D., Maguire, E.A., 2005. The human hippocampus: cognitive maps or relational memory? J. Neurosci. 25 (31), 7254–7259.
Leslie, A.M., 1992. Pretense, autism, and the theory-of-mind module. Curr. Dir. Psychol. Sci. 1 (1), 18–21.
Leslie, A.M., 1994. Pretending and believing: issues in the theory of ToMM. Cognition 50 (1–3), 211–238.
Leslie, A.M., Friedman, O., German, T.P., 2004. Core mechanisms in "theory of mind". Trends Cogn. Sci. 8 (12), 528–533.
Leyens, J.-P., Paladino, P.M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., et al., 2000. The emotional side of prejudice: the attribution of secondary emotions to ingroups and outgroups. Personal. Soc. Psychol. Rev. 4 (2), 186–197.
Macrae, C.N., Bodenhausen, G.V., Milne, A.B., 1994. Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. J. Pers. Soc. Psychol. 66, 37–47.
Malle, B.F., 2005. Three puzzles of mindreading. In: Malle, B.F., Hodges, S.D. (Eds.), Other Minds: How Humans Bridge The Divide Between Self and Other. Guilford Press, New York, pp. 26–43.
Marr, D., 1982. Vision. W.H. Freeman, San Francisco, CA.
McArthur, L.Z., Baron, R.M., 1983. Toward an ecological theory of social perception. Psychol. Rev. 90 (3), 215–238.
McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proc. Natl. Acad. Sci. U. S. A. 98 (20), 11832–11835.
Meltzoff, A.N., Brooks, R., 2001. "Like me" as a building block for understanding other minds: bodily acts, attention, and intention. In: Malle, B.F., Moses, L.J., Baldwin, D.A. (Eds.), Intentions and Intentionality: Foundations of Social Cognition. MIT Press, Cambridge, MA.
Mitchell, J.P., Heatherton, T.F., Macrae, C.N., 2002. Distinct neural systems subserve person and object knowledge. Proc. Natl. Acad. Sci. U. S. A. 99, 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. J. Neurosci. 24 (21), 4912–4917.
Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005a. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. NeuroImage 28 (4), 757–762.
Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005b. The link between social cognition and self-referential thought in the medial prefrontal cortex. J. Cogn. Neurosci. 17 (8), 1306–1315.
Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2005c. Forming impressions of people versus inanimate objects: social–cognitive processing in the medial prefrontal cortex. NeuroImage 26, 251–257.
Mitchell, J.P., Macrae, C.N., and Banaji, M.R., submitted for publication. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others.
Newen, A., Vogeley, K., 2003. Self-representation: searching for a neural signature of self-consciousness. Conscious. Cogn. 12 (4), 529–543.
Nickerson, R., 1999. How we know – and sometimes misjudge – what other know: inputting one's own knowledge to others. Psychol. Bull. 125, 737–759.
Paladino, M.-P., Leyens, J.-P., Rodriguez, R., Rodriguez, A., Gaunt, R., Demoulin, S., 2002. Differential association of uniquely and non uniquely human emotions with the

ingroup and the outgroup. Group Process. Intergroup Relat. 5 (2), 105–117.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 676–682.
Saxe, R., 2005. Against simulation: the argument from error. Trends Cogn. Sci. 9 (4), 174–179.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: fMRI investigations of theory of mind. NeuroImage 19, 1835–1842.
Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43 (10), 1391–1399.
Saxe, R., Carey, S., Kanwisher, N., 2004. Understanding other minds: linking developmental psychology and functional neuroimaging. Annu. Rev. Psychol. 55, 87–124.
Shulman, G.L., Fiez, J.A., Corbetta, M., Buckner, R.L., Miezen, F.M., Raichle, M.E., et al., 1997. Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. J. Cogn. Neurosci. 9, 648–663.
Vogeley, K., Fink, G.R., 2003. Neural correlates of the first-person-perspective. Trends Cogn. Sci. 7 (1), 38–42.