How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition

Ralph Adolphs

Division of Humanities and Social Sciences, California Institute of Technology, HSS 228-77, Caltech, Pasadena, CA 91125, USA

ARTICLE INFO

Article history:
Accepted 30 December 2005
Available online 28 February 2006

Keywords:
Social cognition
Emotion
Simulation
Theory of mind

ABSTRACT

In what ways, and to what extent, is social cognition distinguished from cognition in general? And how do data from cognitive neuroscience speak to this question? I review recent findings that argue social cognition may indeed be specialized, and at multiple levels. One particularly interesting respect in which social cognition differs from the rest of cognition is in its close interaction with the social environment. We actively probe other people in order to make inferences about what is going on in their minds (e.g., by asking them questions, and directing our gaze onto them), and we use the minds of other people as a collective resource. Experiments from our own laboratory point to the amygdala as one structure that is critically involved in such processes.

1. Introduction: beyond building models of the social world

Brains transform an array of sensory information from external stimuli into behavioral responses adapted to interact with those stimuli. The mechanisms whereby they do so is the big question that systems, behavioral, and cognitive neuroscience has chosen to tackle. What is clear is that, for almost all interestingly complex behavior, there is not a simple mapping of the stimulus representation onto the motor representation. Rather, the process is creative and inferential in nature.

Nowhere is this distinction more obvious than in the behaviors towards socially relevant stimuli. Humans and other animals guide their social behavior based on a vast canvas of spatial and temporal context in which a stimulus occurs. The way in which social stimuli link to behavior is highly flexible, and often quite unpredictable. And, notably, humans also make inferences that go far beyond the appearance of the stimuli—inferences about what goes on in the minds and bodies of the people whose actions they observe. How are we able to make such inferences—often fast and reliable—and how is it that they can carry so much conviction that there are other minds behind the faces of people we observe, and so potently motivate us to act?

Broadly speaking, the mechanisms that permit social cognition, that give rise to our ability to infer the mental states of others, depend both on processing that can be described as filtering (the elimination of information present in the environment) and on processing best described as creative (the generation of information not present as such in the environment). We filter social information so as to preferentially process that which is deemed most salient, and we construct from it a rich model of the social world that goes well beyond what the senses alone could provide for us. This picture of the senses as filtering the plethora of information with which we are continuously bombarded has a long history, and is perhaps clearest in the idea that the filters are matched to the relevant signals they are designed to detect. The idea that relatively sparse proximal sensory information is then used to reconstruct a model of the distal properties of stimuli

© 2006 Elsevier B.V. All rights reserved.
has a long tradition, and has been worked out perhaps in the
greatest detail in computational theories of vision that took
their initial inspiration from the framework of Marr (1982).

However, as I will suggest in this paper, this view should
not be taken as an exhaustive description of social cognition.
Indeed, its overemphasis may be quite misleading in that it
suggests that essentially all the real work in generating social
knowledge goes on within our brains as they construct these
rich models of the social world based on relatively
impovertished sensory input. Much of what we do may well
be described as model construction and information filtering.
But the generative nature of cognition is driven not only by
inferences we make once sensory information has been
perceived, it is driven also by the possibility of discovering
new information in the environment in the first place. We
explore our environment, and we actively seek out social
information, an issue that has, I think, not been fully
appreciated in social neuroscience and that is ripe for further
investigation.

The point can be brought out by a recent example from
work in our laboratory: the role of the amygdala in recognizing
fear from facial expressions. Our earlier work, and consider-
able work by others using both the lesion method as well as
functional imaging, had suggested the view that the amygdala
was “specialized” for processing information about fear.
Amygdala lesions resulted in a disproportionate impairment
in the recognition of fear from facial expressions, compared to
other emotions (Adolphs et al., 1994; Calder et al., 1996), and
viewing facial expressions of fear resulted in amygdala
activation in healthy individuals (Morris et al., 1996; Whalen
et al., 2001). There were rapidly accumulating exceptions to
this specificity, already an indication that the story was not
going to be so simple (Adolphs, 1999; Kim et al., 2004; Whalen,
1999; Williams et al., 2005). Nonetheless, one interpretation
of the data, and one that I believe is still likely to be part of
the story, albeit not the whole story, ran roughly as follows. When
we initially perceive the face, temporal visual cortices
construct a detailed representation of the stimulus—the
features of the face and their relative configuration. The
filtering component mentioned above would come into play
here. This representation of the visual stimulus, in turn, would
then be associated by the amygdala with a representation of a
somatosensory cortex, a state of the body that would correspond in
some part to the presumed body state of the person whose
face we are seeing. The association triggered by the amygdala
in this way could then simulate within the viewer aspects of
the emotional state inferred about the person shown in the
stimulus. Somatosensory cortices, insula, and other somatic
mapping structures would then represent this emotional
state and provide explicit information about the emotion
(Adolphs, 2002). Thus, the amygdala would link two kinds of
representations: a visual representation of the other person’s
face we are viewing, and a somatic representation that would
represent the presumed emotional state of that person. This
link effect ed by the amygdala could be fairly direct (via direct
projections from the amygdala to the insula, an interoceptive
somatosensory cortex), or more indirect (via first eliciting an
actual emotional response in the viewer’s body that could
then subsequently be represented in structures like the
insula).

This account of how we might infer another’s emotional
state via simulation (Goldman and Sripada, 2005) is quite
consistent with the model-building picture sketched above.
However, it turned out to be an incomplete picture. A more
recent study gave the surprising finding that the amygdala
comes into play in a more abstract, and earlier, processing
component (Adolphs et al., 2005). Amygdala damage was
found to impair the ability to use information from a
diagnostic facial feature—the eye region of the face. Following
amygdala damage, the eye region of faces was no longer used
effectively by the viewer in order to discriminate fear. These
findings were consistent with other results showing amygdala
activation to fearful eyes (Morris et al., 2002), or only to the
briefly presented whites of eyes (Whalen et al., 2004). In fact,
the deficit was even more basic than that: the reason that
information about the eye region was not used effectively in
viewers with damage to the amygdala was because the eye
region was not fixated by them in the first place. In a final
experiment, we instructed a patient with bilateral amygdala
damage to direct her gaze onto the eyes of other people’s
faces, and found that this manipulation temporarily allowed
her to generate a normal performance on a fear recognition
task in which she was otherwise severely impaired.

It is worth noting two key further results from this study.
First, the subject with amygdala lesions failed to fixate the
eyes in any face, not just facial expressions of fear. In fact, she
simply failed to explore faces in general, which included a
failure to direct her gaze towards the eye region. Similarly,
the abnormal use of information from the eye region held for
happy faces as well as for fearful faces. So the impairment in
use of information from, and fixation onto, the eyes in faces
was general for faces. The reason that this general impairment
resulted in a relatively specific impairment in fear recognition
was just because the eye region of the face was in fact the most
diagnostic for signaling fear, rather than other emotions.
Given the recognition tasks we used, this resulted in a severe
impairment in recognizing fear, but not in recognizing other
emotions. (Interestingly, unpublished data indicate that the
same subject does fixate the eye region when the faces are
shown inverted. So, while the brain does not need to know
that the face is showing fear in order for the impaired eye
fixations to occur, it apparently does need to know that the
stimulus is a face; the impairment in fixation does not seem to
generalize to objects other than faces, including inverted
faces.)

A second point worth noting is that the explicit instruction
to fixate the eyes in faces, while rescuing the subject’s
impaired recognition of fear, did so only transiently (as long
as that block of the experiment lasted). When later asked to
view faces, the subject spontaneously reverted to her lack of
exploration of the face, and once again showed impaired fear
recognition. One reason that the improvement was not more
permanent may well be that the patient was not given
additional information about her impairment. She was
unaware that she failed to fixate the eyes, as she was unaware
that her performance in fear recognition was impaired. This
raises further questions: why did she not ask about her
performance, why did she not notice that she failed to fixate
the eyes? I believe that these questions point towards a
broader interpretation of the impairment, and one that is in
line with the role for the amygdala I am sketching in this paper. The subject, as a result of damage to the amygdala, lacked a normal mechanism to explore the environment. One aspect of such an impairment was a failure to fixate the eyes in faces, to explore them normally with one’s gaze. Another aspect of the impairment was a failure to question what was going on in the experiment in any way, or to monitor one’s own performance during it. In both cases, there remains a passive ability to process sensory information, but the instrumental component of seeking out such information in the first place has been severely compromised.

These new data indicate that the amygdala comes into play much earlier than initially thought, and in a more abstract way that is not specialized for recognizing fear as such. It appears to be specialized for seeking out potentially salient social information in the first place, by directing our gaze and visual attention to certain regions of faces that should be explored in more detail. It may be that this role extends beyond faces to a broader role in exploration of the social environment generally (Sander et al., 2003), as the above discussion suggests, similar to earlier proposals that the amygdala serves to detect potentially important stimuli about which more information must be gathered (Whalen, 1999).

The above example sets the stage for the argument I will be defending in this paper: that social cognition involves loops of processing that are extra-neural. It involves the bodies, and the social environment, in which brains are embedded. The philosopher John Dewey already emphasized this enactive aspect of cognition when he wrote.

“Upon analysis, we find that we begin not with a sensory stimulus, but with a sensori-motor coordination, the optical-ocular, and that in a certain sense it is the movement which is primary, and the sensation which is secondary, the movement of body, head and eye muscles determining the quality of what is experienced. In other words, the real beginning is with the act of seeing; it is looking, and not a sensation of light. The sensory quale gives the value of the act, just as the movement furnishes its mechanism and control, but both sensation and movement lie inside, not outside the act” (Dewey, 1896). In this paper, I propose to take this view seriously, and to explore some of the consequences it has for social neuroscience.

The new findings described above (Adolphs et al., 2005) add a new component that must be considered. Social cognition draws upon mechanisms that show some degree of specialization in terms of perceptual processing of sensory stimuli, and in terms of the kinds of internal models that are constructed to permit inferences from those stimuli. Domain-specific processing of faces and empathizing with others are two examples of these features. But a third component needs to be added: mechanisms for exploring the social environment and for probing it interactively. Taken together, these different components of social cognition suggest that our knowledge of the internal bodily and mental states of other people derives from detecting specific features, making specific inferences, and asking specific questions of the social environment (not at all in that order but in all likelihood concurrently). Perhaps what distinguishes social cognition from cognition in general is the extent to which these three components are integrated, and especially the extent to which the first and third are connected in the reciprocal social interactions that we typically engage.

2. Is perception of social stimuli special?

The first two of the three features of social cognition outlined above have typically been claimed to show that social cognition is “special”, in the sense that it draws on psychological processes, and on neural structures, that other aspects of (nonsocial) cognition do not (or do to a lesser extent). Perception of socially relevant stimuli, best studied in the visual modality, and making inferences about the mental states of other people, as probed for example with “theory of mind” tasks, all share features of domain-specificity.

Let us take the perceptual component first. The broad approach to identifying classes of sensory stimuli that might be of social significance to a species is essentially what neuroethology aims to achieve: observe the behaviors of animals, in their natural environment, to naturalistic stimuli. This approach has certainly shown that there is remarkable specificity, insofar as the behavioral responses of animals are exquisitely tuned to very specific stimulus parameters that have ecological meaning (Ghazanfar and Santos, 2004), and several studies have now begun to elucidate the neural substrates of such processing (Gil-da-Costa et al., 2004). Most of the research in humans has focused on stimuli that we all know from firsthand experience have social meaning, and the largest corpus of studies has explored the processing of faces. There is evidence that regions of higher-order visual cortex are disproportionately engaged by faces or by biological motion, as compared to other classes of visual stimuli. There is now an extensive set of studies, primarily from functional neuroimaging, that document the differential activations of certain occipital and temporal visual cortices to viewing bodies (Downing et al., 2001), biological motion (Grossman and Blake, 2002), and faces (Kanwisher et al., 1997), and supporting evidence that the perception of these classes of social stimuli are dissociable from lesion studies (Duchaine et al., 2004; Heberlein et al., 2004; Moscovitch et al., 1997).

Data such as these bolster the view that our processing of social stimuli is already specialized at the level of perception, a question that has been debated the most in regard to the perception of faces. Behaviorally, there is evidence that faces are attended preferentially very early in life, that they are processed in a configural way such that relations between their features are encoded (but only for upright faces), and of course we have the everyday observation that we are simply all experts at recognizing people from their faces. Brain imaging data have documented a mosaic of regions of visually responsive cortex in the temporal and occipital lobes that is activated disproportionately for faces as compared to other visual stimuli (Haxby et al., 2001; Spiridon and Kanwisher, 2002). One region in particular, the “fusiform face area”, appears to distinguish more between faces than any other visual objects (Spiridon and Kanwisher, 2002). These data have supported the well-known view that there is domain-specific processing of faces (Kanwisher, 2000; Kanwisher et al., 1997). As is also well known, this view has not gone unchallenged. Based in good part on data from artificial
stimuli called "greebles", it has been argued that these regions of visual cortex reflect specializations for executing certain processing strategies, rather than for certain stimulus classes (Tarr and Gauthier, 2000). There is evidence going both ways—activation studies do suggest that the fusiform face area can be activated by greebles as well as faces, although whether to the same degree remains an open question. Lesions typically impair more than just face recognition as such, although there may be rare cases highly specific to only a face processing impairment (Duchaine et al., 2004).

There is the second domain where there is evidence that social cognition is special. That is the domain of "mentalizing", of inferring other minds and their states and contents from observations of people (Blakemore and Decety, 2001; Frith and Frith, 1999). It remains a heated debate whether any other primates might have similar capacities, or precursors to them (Povinelli and Vonk, 2003; Premack and Woodruff, 1978; Tomasello et al., 2003), but it seems clear that this kind of inference does not apply to non-social stimuli.

The sets of perceptual cues that trigger mental state attributions can be quite impoverished, as reflected in our ability to make social inferences from very brief slices of dynamic visual stimuli (Ambady and Rosenthal, 1992) and our propensity to anthropomorphize even stimuli that are clearly not inherently social. A classic example has been the visual motion of geometric shapes, studies pioneered by Johansson, Michotte, and Heider (Heider and Simmel, 1944; Johansson, 1973; Michotte, 1946). Heider’s work in particular stimulated the design of various animated stimuli in which simple shapes (triangles, circles) move so as to convey emotions and other social descriptions (Castelli et al., 2000). Healthy subjects automatically make such attributions, and in a sense perceive a "social illusion" from these stimuli. Interestingly, viewing the stimulus appears to recruit some of the same visual cortical regions that are activated when we view faces (Schultz et al., 2003), and the social interpretation of the stimulus is abolished or reduced in people with autism (Abell et al., 2000; Klin, 2000) or in neurological subjects with bilateral amygdala damage (Heberlein and Adolphs, 2004). These latter data prompted a further question: what is driving the activation of the face-specific visual regions in this case? Is it that stimuli such as moving geometric shapes share some features in common with faces? Or, as would seem more likely, is it that higher-order processes influence our very perception of the stimuli?

This question is not as murky as it seems. Clearly, there are certain aspects of information processing that are very close to the stimuli, so to speak (e.g., processing within the retina), whereas others are relatively distal to the stimuli and closer to what we say about them (e.g., neural systems for language production). While there is no black-and-white dichotomy between percept and concept in neural terms, there is a graded distinction. At the level of psychology, folk or otherwise, there is the distinction between those aspects of the mind that are sensory in nature (e.g., our visual experience of how a person’s face appears to us) and those that are conceptual in nature (e.g., what we know or judge about the person whose face we see). The former aspects are usefully probed with tasks such as detection and discrimination, whereas the latter ask people to tell us what they believe about what they see.

I think that many people would count specializations at the sensory side as the stronger evidence than specializations on the conceptual side for the claim that social cognition is “special” in some way. The debate noted above regarding the putative specialization of the fusiform gyrus for processing faces, for example, appears aimed at this level. The claim is not that we employ different concepts in thinking about faces or different words in talking about them, than would be the case for non-social stimuli other than faces. That seems like a trivially true, and relatively uninteresting fact. Rather, the claim is that the very reason that we think and talk differently about faces than about other visual objects is that we perceive them differently in the first place.

So, what evidence is there? It seems to me there is very good evidence indeed that some social stimuli are processed “specially”, and the evidence is there at all levels of processing. At the level of transduction and very early sensory processing, detection of pheromones would seem to be an ideal example of a highly specialized aspect of social information processing in invertebrates (Krieger and Ross, 2002) as well as mammals (Lin et al., 2005; Stern and McClintock, 1998). Likewise, the evidence of face-selective neurons in the temporal lobe of primates (Perrett et al., 1987), and song-selective neurons in the forebrain of songbirds (Margoliash, 1986) look like strong examples of sensory processing that is highly specialized for social stimuli. Whether this is set up through experience with a world that happens to contain, and make requirements about the discrimination among, those social stimuli, or whether it is independent of such experience is an important further question, but not one that we need to answer here.

But there is evidence at the level of concepts as well. We think about people in a way that we do not, and probably cannot, think about inanimate objects: we accord them points of view on the world, are concerned about what they think of us, and give them moral rights that chairs and tables cannot have. The latter may be especially definitive of a conceptual-level specialization for social cognition. Thinking about people has a normative character that thinking about non-social stimuli does not. In fact, even thinking about nonhuman animals does not seem to engage the same moral attributions that we accord to people.

A well-known example of specialization at the level of judgment and reasoning are the content-specific effects in the Wason selection task (Wason and Johnson-Laird, 1972). Not
only is there evidence (albeit, as with all examples in this section, debated) that syllogistic reasoning involving social exchange can be specialized for the detection of violations of social contracts (i.e., detecting people who cheat) (Cosmides and Tooby, 1992), but there is even neurological evidence suggesting neural substrates for the effect (Adolphs et al., 1995; Stone et al., 2002). The interpretation of these findings is typically that the content of what we think about causes us to use a certain scheme in thinking about it: specific contents engage specific processes.

A possible neural substrate for reasoning about social material is the prefrontal cortex. Lesions of the ventromedial prefrontal cortex (as well as to other structures involved in regulating somatic states) result in a disproportionate impairment in emotional and social intelligence, compared to the usual cognitive intelligence (Bar-On et al., 2003). And there are of course the well-known studies implicating regions of prefrontal cortex in social behavior (Damasio, 1994) and theory of mind (Gallagher and Frith, 2003). Just as with the critiques regarding the specialization of sensory cortices for processing specific to the social domain, as we saw in the case of the fusiform face area, there are worries about whether some of the more central cognitive processes just discussed capture the “social” nature of social cognition, or whether they might share something else in common that is not necessarily “social” as such, but merely comes into play typically but not exclusively when we process social stimuli. For instance, perhaps mentalizing, social intelligence, and related competences are measured by tasks that are also more complex, or require more flexible reasoning that typically do tasks designed to assess our reasoning about the nonsocial world. If this was the case, by analogy with the debate regarding the domain-specificity of face processing, we would have shown that social stimuli make certain processing demands, but not that there are psychological or neurological mechanisms that are necessarily specialized to process social stimuli. For instance, the medial prefrontal and cingulate cortex involvement often seen in theory of mind tasks (Gallagher and Frith, 2003; Gallagher et al., 2000, 2002; Stuss et al., 2001) might result from increased cognitive effort and control needed to perform those tasks, as these factors are also known to recruit this region of the brain (Ridderinkhof et al., 2004; Williams et al., 2004). The role of orbitofrontal cortex in social cognition may derive from its role in counterfactual thinking (Camille et al., 2004). But this does not mean that neural systems involving regions like medial prefrontal cortex are not adaptations for social cognition: it may well have been precisely our complex social nature that required factors like increased cognitive control (Barrett et al., 2003).

Even in cases where there is reason to think that the social cognitive functions of a brain region are derivative to more general, nonsocial functions, we do not have to jump to the conclusion that therefore there is no “specialization” for social cognition. It may be that social cognition is best thought of as a particular mode of processing, one that can modulate cognition in multiple domains including perception, attention, memory, and decision-making (Mitchell et al., 2002, 2004).

In order for differential processing of socially relevant stimuli, there have to be processing mechanisms in place in the brain that are engaged differentially when they are fed social stimuli as opposed to non-social stimuli. But what distinguishes social from non-social stimuli, of course, has also to be inherent in systematic differences in the stimuli on the basis of which such differential neural discriminations could be made. Is the social nature of stimuli confounded by their non-social features? This question is analogous to the above question of whether social cognition is confounded by all the other aspects of cognition that comprise it. Consider one study from our own laboratory, in which we found that neurons in the prefrontal cortex responded very rapidly to a semantically specified stimulus category (Kawasaki et al., 2001). We found that these neurons showed differential responses, with a short latency of around 120 ms, to visual stimuli that were judged to be aversive, as opposed to visual stimuli judged to look pleasant or neutral. Now, on what basis did this differential neural response occur? One possibility that would be interpreted as a “confound” would be if the aversive visual stimuli were also distinguished on some other, low-level visual property. For instance, had it been the case that the aversive images were all brighter or larger than neutral and pleasant stimuli, then the differential neural response we found might have been discriminating brightness or size rather than emotion category. We checked, and it was not the case that the aversive stimuli were simply brighter or larger. Moreover, they were extremely heterogeneous (pictures of snakes, spiders, mutilations, war, etc.) making it very unlikely that they shared in common any simple feature that could explain the neural discrimination. So our interpretation of the finding was that the neurons were encoding how aversive human viewers judged them to be: the emotion category was created by the brain rather than inherent in the stimuli. We did not do the further experiment, but the prediction would be that individual differences in how viewers would judge the emotion of the stimuli, for the same set of stimuli, would correlate with the neural responses. On the other hand, it seems obvious that the emotion discrimination has ultimately to be based also on the features of the stimuli—just not rigidly so.

### 4. Simulation and empathy

I want now to discuss a particular mechanism for making inferences about other people that is often thought to be “special” to social cognition: our ability to conceive of others like ourselves, and to obtain knowledge about other people by imagining what it would be like to be them. Originally articulated by Titchener, Lipps and others (Lipps, 1907), this idea has received considerable recent attention with the discovery of “mirror neurons” that respond both to one’s own actions as well as to those of a conspecific, and the discovery that somatic mapping structures in the brain are activated both when we feel an emotion and observe another person express it (Blakemore and Decety, 2001; Gallese, 2003; Gallese and Goldman, 1999; Rizzolatti et al., 2001). (It should be noted that these findings do not necessarily point to the same theory; in fact, there are several competing versions of simulation theory).

There are now several studies indicating that the observation of another person’s emotional state recruits structures...
like the insula (Jackson et al., 2005; Singer et al., 2004), an interoceptive somatosensory cortex also involved in representing our own somatic states. Interestingly, the insula has been hypothesized (Craig, 2002; Damasio, 1999) and recently shown (Critchley et al., 2004) to be associated with the conscious experience of our own body state. This suggests that our knowledge of another person’s emotional state through simulation of their presumed somatic state relies on a simulation that is explicit, in the sense of providing conscious access to the emotion that is being simulated (although not necessarily awareness of the fact that the perception of the other person caused the emotion). That is, the simulation mechanism through which we infer another person’s emotion is empathic: it involves actually feeling (aspects of) the emotion of the other person. In this sense, the output of knowledge by simulation may be quite different than the output of knowledge by reasoning alone: it is a distinction between knowledge by acquaintance versus knowledge by description.

This ability to generate knowledge via constructing a model that provides a particular conscious experience as the output is, however, not unique to social cognition. Any kind of imagery essentially achieves the same thing, albeit not necessarily via creating a somatic image. The advantage is similar in all cases: creating a visual image of your house in order to answer the question, “how many windows are there in your house?” provides fine-grained information that is simply not available via retrieval from semantic memory; creating a somatic image of the emotional state of another person to answer the question, “how does this person feel?” provides fine-grained information about their internal state in the same way. By making the output of the simulation a consciously accessible sensory image, it is afforded the flexibility needed to actively explore it in much the same way that would explore the actual stimulus.

Several further questions arise about the nature of the simulations we use to mentalize. To what grain do we simulate? How do we distinguish the simulation from the real thing? Again, the answers are probably quite parallel to those we would give in regard to visual imagery: the grain likely depends on the level of knowledge that is to be reconstructed (whether it is at a superordinate or subordinate level of categorization); the simulation is distinguished from the real thing both by the fact that the two representations are not entirely overlapping (Jackson et al., 2005; Keysers et al., 2004; Singer et al., 2004) and by the presence of additional structures that allow us to distinguish the two (Ruby and Decety, 2001).

Although somewhat peripheral to the present paper, it should be noted briefly here that the sensorimotor simulation theory is not without its detractors. However, the critiques focus on specific and more restricted examples of the theory than the general account I have sketched above. As such, they are well taken; but they do not demonstrate flaws in the general idea. For instance, one criticism (Jacob and Jeannerod, 2005) is leveled against the idea that motor simulation could account for inferences of unobservable states like beliefs if mere imitation of observed actions is the mechanism. Relatedly, there is a worry about how simulation of another’s action would conflict with premotor planning of one’s own (e.g., when seeing an angry person, one would need to simulate the other’s anger in addition to having one’s own fear). The answer to both charges seems to me to be that mirror neurons are not the whole story. A sensorimotor system of several structures will underlie simulation, and different partitions of the representation of body states or action plans will be used to model one’s own versus those of another person (or several people that are being observed, for that matter). A second critique (Saxe, 2005) also focuses on the “mirroring” aspect of simulation that has been suggested by the discovery of mirror neurons, arguing that the patterns of errors subjects make on tasks are inconsistent with such an account. The answer to this critique, like the first one, is again that indeed mirror neurons are implausible as the sole substrate of simulation and, more broadly, that simulation is implausible as the only strategy for mentalizing. No doubt we do use more “theory” oriented reasoning strategies to make inferences about others’ minds, and simulation is not the only game in town. But that does not show that simulation, broadly conceived, is not a good part of the story.

In connection with the above critiques, it is also worth pointing out that, in fact, we routinely encounter situations in which we have to make inferences about other people’s bodily and mental states that are not based on cues that would directly lend themselves as input to a mirroring system. For instance, we hear through third-person accounts, or read in a novel, what a person might be doing, rather than find out through direct observation of the person. It is possible that mirror neurons as such play no role in inferences about other minds based on such information; rather, self-referential thinking, perhaps without the engagement of emotional or motoric components, may suffice in such cases. Of course, it is also possible that mirror neurons do play a role, albeit based on input from already inferential processes that have constructed images based on the information that was provided. It seems to me that different simulation accounts, or mentalizing accounts that are not simulationist, are almost never the entire story and almost never mutually exclusive in telling the whole story. They may all come into play, at various times, and depending on task demands.

The above noted parallels with our imagery-based models of the nonsocial environment notwithstanding, there are two features of simulating other people that seem unique: the link to action and the body. When we feel another’s emotion through our simulation of them, we also feel the urge to act on the basis of that emotion (deGelder et al., 2004). Insofar as emotions are intrinsically motivating, our empathic responses to other people immediately motivate us to act—for instance, to help the other person or to avoid them. This point, that our knowledge of other minds is tightly linked to our interactions with others is, I think, an important distinction of social cognition and one to which I will return at the end. Our perception of other people is closely tied to our concern for them and our propensity to behave with respect to them. This is also, in my view, the reason that moral judgments about other people are immediately linked to a motivation to behave to help or punish them. Here, as well, emotion comes into play in motivating us to act on our judgments.

The role of the body also makes simulation different from other forms of imagery. It may not be necessary for an overt
somatic response to mediate between perception of others and our simulation of their internal states (Damasio, 1994; Heims et al., 2004). Possibly, structures that could trigger actual responses in the body, such as premotor cortices or the amygdala, could also trigger changes more directly in the neural representations of body states, such as those in the insula (Carr et al., 2003). Nonetheless, we do in fact often engage the body in simulating other people, and it remains possible that we always engage it to some extent. For instance, observing other people express emotions results in some mirroring of the physiological emotional state in the viewer (Dimberg, 1982). The possibility of using the body itself as the substrate of the simulation when we model another person’s emotion would be not only economical, but suggests an interesting way in which actual, analog physical processes—state changes in various parameters of the body that normally comprise an emotional response—can be used in information processing. The body might be thought of as a somatic scratchpad that we can probe with efferent signals in order to reconstruct knowledge about the details of an emotional state. Given the complexity of interaction among multiple somatic parameters, in action as well as emotion, it may not be feasible to carry out the same simulation entirely neurally.

There is a final feature of body-based simulation to which I will return at the end: by engaging our own bodies in simulating those of another person, we also express a social signal that can, in turn, be perceived by the other person (Adolphs, 2001). This closed loop between perceiver and observer in many social interactions highlights what Darwin had already noted about facial expressions: their social communicative nature (Darwin, 1872/1965).

5. Microscopic and macroscopic specialization

The question of whether social cognition is special applies at levels of description ranging from genes to behavior. As with the stages of processing discussed above, our reductionist predilection often tends to view evidence at the more microscopic level as stronger than evidence at the more macroscopic. Good examples are the finding that gene knockouts can disproportionately affect aspects of social behavior. Oxytocin-knockout mice, for instance, show impairments in memory that are selective for the memory of the odors of the conspecifics (Ferguson et al., 2000), in line with other evidence linking this peptide to social behavior in rodents (Insel and Young, 2001; Young and Wang, 2004) as in humans (Kosfeld et al., 2005). Opioid-receptor knockout mice show impaired attachment behavior between mother and pups (Moles et al., 2005). However, these microscopic levels of description should not be given any more weight than more macroscopic or systems-level ones. Asking whether there is “a gene”, “a neurotransmitter”, or “a brain structure” for social cognition, or for a particular emotion, is the wrong question to ask, because it presupposes that a single level of description is the only appropriate one.

A good example of how different levels of description interact comes from recent findings on affective disorders. Take the example of depression. There are now findings implicating genetic polymorphisms in the serotonin reuptake transporter promoter to the risk for depression (Caspi et al., 2003) and to altered function of specific brain structures like the amygdala (Hariri et al., 2002). Notably, the altered function is more complex than changes in the activity within a single structure: a genetic predisposition for traits correlating with risk for depression was associated with differential functional connectivity between amygdala and cingulate cortex (Pezawas et al., 2005). Another finding is that the actions of antidepressants apparently require neurogenesis in the hippocampus, because blocking neurogenesis also blocks the behavioral effects of antidepressant drugs, a puzzling interaction between the cellular and molecular level (Santarelli et al., 2003). A further interaction is provided by the finding that hippocampal neurogenesis is modulated (increased) by the social status of the animal (Kozorovitskiy and Gould, 2004). An increasing number of studies are now taking into account individual differences in personality traits as well as in spontaneous as well as volitionally regulated emotional state to account for differences in brain function (Davidson et al., 2000). For instance, activation in the amygdala to emotional facial expressions is modulated based on the trait (Etkin et al., 2004) or state (Bishop et al., 2004) anxiety of the subject, and instructed modulation of emotional experience influences the amygdala response to emotional stimuli (Schaer et al., 2002).

The prefrontal cortex also provides a good example of evidence for specialization of sorts at multiple levels. The volume of frontal cortex appears to have expanded, relative to the rest of the brain, in primates (Bush and Allman, 2004), although humans do not appear distinctive in this regard compared to other apes (Semendeferi et al., 2002). There is some suggestion that more anterior regions of frontal cortex may indeed be different in apes compared to other primates, or perhaps humans compared to other apes; whether this is due to changes in grey matter volume (Semendeferi et al., 2001) or increased connectivity (Schoenemann et al., 2005) remains unclear (probably both are important).

In addition to these volumetric data, there is evidence that pyramidal cells in the prefrontal cortex of primates and humans are distinguished morphologically (Elston et al., 2001), perhaps reflecting the differential roles played in higher cognitive functions that also contribute to social behavior. An even more striking example are the Von Economo cells of anterior cingulate and frontoinsular cortex, large spindle-shaped neurons that are unique to humans and great apes (Nimchinsky et al., 1999) and have been hypothesized to function in social emotions (Allman et al., 2005).

What this somewhat bewildering brief tour through different levels of description suggests to me is that the kinds of neurobiological accounts that we will ultimately give of social cognition are likely to cut across multiple levels. This is already so in several of the papers cited above. While we can look for, and to some degree find, specializations for social cognition at each level taken individually, the challenge eventually will be to come up with an account that relates several different levels of description and that explains the relations between them that result in specialization.
6. Challenges for the future

I want to end by considering three issues that I think are important challenges for future studies in social cognition. The first is methodological, but I think related to the other two. This is the issue of ecological validity. Essentially, all neuroscience data on social cognition come from stimuli that are “social” only in a highly derivative way. Typical examples are static photographs of facial expressions. Participants in the experiments know full well that these are not real people, and although many aspects, especially of perceptual processing, may be shared in common between such stimuli and the real thing, they clearly lack the interactive and meaningful nature that a real person would provide.

There is no need to dwell on this issue because it is universally acknowledged, and because it is in fact now being surmounted. A good example are interactive experiments in which two participants have to make strategic choices, often to win or lose money in “games” of the sort that behavioral economists have studied (Camerer, 2003) (or nonhuman primate versions of these (Barracough et al., 2004)). These protocols have recently been translated into the fMRI environment, and not only let us examine the neural correlate of a real social interaction (deQuervain et al., 2005; Gallager et al., 2002; Shergill et al., 2003), but also provide the opportunity in future studies to analyze neural activity in the brains of both players as a coupled system (King-Casas et al., 2005).

A second challenge is to explicate how social cognition relates to the distinction between conscious and nonconscious processing. This question is very related to the question of whether social cognition is “special”: evidence that it is special at the level of automatic, implicit processing is counted more heavily by most people than evidence that it is special at the level of conscious, volitional processing. Neural responses to faces that cannot be consciously perceived because of brain damage (Pegna et al., 2005) or subliminal presentation (Morris et al., 1999; Whalen et al., 2004), implicit biases towards person categories revealed with the implicit association test (Greenwald and Banaji, 1995), and preferences for people based on cues of which they are unaware all tend to make an impression, and much of social psychology has focused on the influences of memory schemas on our social judgment and behavior that lies outside of our conscious awareness (Ferguson and Bargh, 2004).

Yet, as we noted above, the final model of another person we construct appears to be typically and largely accessible to consciousness. The nonconscious influences on social behavior that social psychologists have studied so much are, I think, best viewed as the inputs to a model of another person, which is itself something we can access consciously and thus use flexibly. We may often not know why we feel a certain way, or have a certain thought, about another person—but the feeling or thought as such surely seems consciously accessible. The explicit nature of the simulation may also account for another feature of our knowledge of other minds: their indubitability. While we can feel uncertain about the details of what another person is thinking or feeling, it is difficult to doubt that they are feeling and thinking at all. In general, we cannot seriously entertain the skeptical doubt that other people around us do not have minds similar to our own. It is noteworthy that this is so, because we do not seem to have the same difficulty doubting many other inferences we make about the physical, non-social environment. The reason for this asymmetry, and for why we are compelled to take a normative stance when making judgments about the person we observe, could derive from the fact that the mechanism by which we derive these judgments is a consciously accessible simulation. Just as we cannot, in general, doubt our own minds, so we cannot doubt the existence of the minds of others: after all, we literally feel their minds within ours. These considerations may explain why it is quite possible to disbelieve visual illusions, despite their persistence; yet patients with Capgras syndrome hold delusional beliefs about other people and their minds precisely because they fail to have the feelings that a simulation might provide (Ellis and Lewis, 2001; Ellis et al., 1997).

There is a final, important consideration that speaks to the question of what it might be that could be special about social cognition. All of the discussion has focused on mechanisms internal to the individual: mechanisms within the brain, or encompassing the brain and the body. While neural and somatic processes certainly play their role in generating a model of the social world, it is wrong to think of this as exhausting the strategies whereby we find out about other people’s minds. Think of any everyday example in which you are engaged in generating knowledge about what is going on inside someone else—how they are feeling, what they are thinking or intending. You might look at their face, their direction of gaze, and make inferences and run internal simulations based on those visual cues. But you might also walk up to them and ask them, or cast a glance at them to see how they look back at you in return, or smile at them and see if they smile back. That is, we actively probe the social environment in order to glean relevant information. Social cognition is more than just reactive: it is instrumental.

This idea is not news to aficionados of “situated cognition”, who have long maintained that our brains do not store all knowledge about the world in explicit form, and do not hold comprehensive explicit models or representations of the environment. Rather, it has been argued, our brains contain recipes for seeking out that information—often rather trivially by deciding where in the environment to look. The by now classic studies of phenomena such as change blindness seem to show exactly this: we do not form a rich internal visual model that we can inspect, but rather rely on visual inspection of the external world (Noe, 2004; Simons and Rensink, 2005). This idea does not contradict what I said previously about us having a consciously accessible model of other people, it just says that what we access is not necessarily entirely in the brain of the perceiver. Depending on the circumstances (notably the presumed reliability of the internal model versus the sensory evidence), we would rely on an internal simulation or on probing the external social environment. The partition between processes internal and external to the perceiver would need to be flexible to accommodate factors such as speed and reliability that could shift our emphasis from one to the other as suitable to the demands of the situation. Similar ideas are also to be found in social psychology approaches to memory—for instance, work on
transactive memory acknowledges that memory structures often operate in social groups, especially for individuals who have close social ties to one another (e.g., [Wegner et al., 1991]). Indeed, insofar as all of human culture is predicated on the collective cognitive abilities of large social groups, it is more appropriate to view an individual brain not as the repository of social knowledge, but rather as a source for generating it within a supportive social context.

These thoughts suggest that perhaps we should consider social cognition more broadly, as a collection of processes for navigating the social world, that is based not only on events occurring within ourselves. It may encompass the web of social interactions in which we are engaged with other people around us. To find out how they feel and what they think, we probe them, we ask them. In so doing, we do not only find out specifics about other individuals, but we also are able to create a shared space for collective knowledge and expertise. Perhaps, no less importantly, we also find out things about our own minds by relying on the feedback we obtain from other people. In a sense, the social mind is collective, and the “representations” or “models” of social information that are the topic of social neuroscience are only a part of the mechanism by which we know about the minds of others and our own.

Acknowledgments

Supported in part by grants from the National Institute of Mental Health, the Cure Autism Now Foundation, and the Pfeiffer Research Foundation.

REFERENCES


Michotte, A., 1946. La Perception De La Causalite. Institut Superieur de Philosophie, Louvain, France.


