1-1-2008

Direct perception in the intersubjective context

Shaun Gallagher
University of Central Florida

Direct perception in the intersubjective context

Abstract

This paper, in opposition to the standard theories of social cognition found in psychology and cognitive science, defends the idea that direct perception plays an important role in social cognition. The two dominant theories, theory theory (TT) and simulation theory (ST), both posit something more than a perceptual element as necessary for our ability to understand others, i.e., to "mindread" or "mentalize." In contrast, certain phenomenological approaches depend heavily on the concept of perception and the idea that we have a direct perceptual grasp of the other person's intentions, feelings, etc. This paper explains precisely what the notion of direct perception means, offers evidence from developmental studies, and proposes a non-simulationist interpretation of the neuroscience of mirror systems.

Keywords

direct, intersubjective, perception, context

Disciplines

Arts and Humanities | Law

Publication Details

Gallagher, S. 2008, 'Direct perception in the intersubjective context', Consciousness and Cognition, vol. 17, no. 2, pp. 535-543.

Consciousness and Cognition 17 (2008) 535–543

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Target Article

Direct perception in the intersubjective context

Shaun Gallagher *

Philosophy and Cognitive Science, University of Central Florida, Orlando, FL 32816-1352, USA
School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK

ARTICLE INFO

Article history:
Received 28 February 2008
Available online 28 April 2008

Keywords:
Direct perception
Theory theory
Simulation theory
Mirror neurons
Social cognition

ABSTRACT

This paper, in opposition to the standard theories of social cognition found in psychology and cognitive science, defends the idea that direct perception plays an important role in social cognition. The two dominant theories, theory theory (TT) and simulation theory (ST), both posit something more than a perceptual element as necessary for our ability to understand others, i.e., to "mindread" or "mentalize." In contrast, certain phenomenological approaches depend heavily on the concept of perception and the idea that we have a direct perceptual grasp of the other person's intentions, feelings, etc. This paper explains precisely what the notion of direct perception means, offers evidence from developmental studies, and proposes a non-simulationist interpretation of the neuroscience of mirror systems.

1. Introduction

In this paper, I want to defend the role played by direct perception in social cognition. A theory that employs direct perception is quite different from the standard theories of social cognition found in psychology and cognitive science today. The two dominant theories, theory theory (TT) and simulation theory (ST), both posit something more than a perceptual element as necessary for our ability to understand others, i.e., to "mindread" or "mentalize." In contrast, certain phenomenological approaches depend heavily on the concept of perception and the idea that we have a direct perceptual grasp of the other person's intentions, feelings, etc. This concept, however, appears rather suspect from the perspectives of TT and ST, and indeed, in contrast to the great effort mounted by TT and ST to solve the problem of other minds, the direct perception approach comes close to suggesting that there is no problem of other minds. Whoever proposes such an approach, however, is called upon to say precisely what the notion of direct perception means, and to defend this concept against the suspicion that it offers a rather simplistic account of social cognition that fails to get at the real problem.

2. Smart and not-so-smart perception

Perhaps the best way to start, and to provide context for this issue, is to look at the concept of perception as it is used in TT and ST. Both TT an ST start with perception understood as a third-person process, that is, as an observation of the other person, but each adds to perception certain cognitive elements which allow us to understand the other that we observe.

TT contends that the normal and pervasive way in which we understand other persons depends on a practice of mentalizing in which we employ common sense or folk psychological theory about how mental states (beliefs, desires, intentions) inform the behaviors of others. Understanding the other person's mental states allows us to explain or predict the other

* Fax: +1 407 8236658.
E-mail address: gallaghr@mail.ucf.edu

1053-8100/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2008.03.003

person's behavior or what they must be thinking. This theory, or our ability to employ it, may be the product of an innate mechanism, or it may be acquired through early experience, but it initially manifests itself around the age of four years, and this is signaled by the ability of four year olds to pass false-belief tests.[1]

ST claims that we have no need for a theory like this, because we have a model, our own mind, that we can use to simulate the other person's mental states. We begin by observing the other person's behaviors in specific environments, and by simulation we go on to model their beliefs and desires as if we were in their situation.

Note first that for TT and ST, some extra-perceptual cognitive elements seem to be required because of the way the problem is framed. In the standard versions of TT and ST the problem is framed as a problem of access to the other person's mind. The supposition is precisely that the other person's mental states are hidden away and are therefore not accessible to perception. I cannot see into your mind; hence I have to devise some way of inferring what must be there, based on evidence that is provided by perception.

Second, although both TT and ST hold that the process starts with perception, the concept of perception that is at stake is a very specific one.

(1) As we have indicated, perception means third-person observation rather than something that happens in the context of interaction. My role in what may be happening is not included in the scenario. Rather, I tend to stand at the margins of the situation and make observations. In this case there is a disconnection between perception and anything that might involve my own action.
(2) Perhaps because perception plays this observational role, and the important social aspects of cognition are added on, there is no need for TT and ST to provide anything like a theory of perception, and they do not. In this regard, for example, they remain neutral with regard to the question of whether perception is properly considered to be enactive, in the way that has been argued by enactive theorists of perception (e.g., Varela, Thompson, & Rosch, 1991; Noë, 2004). I will return to this point later.
(3) The perception involved seems to be amazingly impoverished or ill-informed about the particular object (i.e., the person) which the subject perceives. If I were to remain with only this perception I would be totally perplexed or at least puzzled about the other person's behavior. I see what the other person does, but until I call forth some theory, or until I run through a simulation routine, I seem not to have any sense of what that person is up to. Perception, then, obviously requires some kind of supplementation. It is not completely impoverished, however. I recognize the other for what she is—another person, rather than, say, some inanimate object that does not have a mind. My perception is smart enough to discriminate between an agent and an object in the surrounding environment. But to go any further, to understand what that agent is up to, I apparently need some extra cognitive tools.

At the same time, we might ask whether the perception involved here is a kind of direct perception in the simple sense that I could say that I directly perceive what the other person is doing. There are two possible answers. The first would suggest that perception is not good enough to really capture the meaning of what the person is doing. What I perceive directly is that the person's hand is reaching to grasp something. Or that the person is walking away from someone else. But perception is not smart enough to recognize that the person is actually getting a drink, or that the person walking away is doing so for a particular reason, for example, because he is angry. To this not-so-smart perception, we need to add some kind of cognitive machinery that will supplement perception with inferences about what the action means; and from this we go onwards to the possibility of attributing reasons or motives for the action. A second possible answer is that the perception is smart enough on its own, without the supplement of inference mechanisms, to deliver some sense that the person is getting a drink or that the person is angry and motivated to walk away. But this cannot be the case on the TT or ST accounts, because that would be to suggest that people wear their intentions and their emotions "on their sleeve" as the saying goes. Rather, as the task is defined, one needs to get into the other person's head and find out what their mental states are.

From these considerations it seems clear that to explain the concept of direct perception involves several related issues. First, how *smart* perception may be—how informed it is. So let us distinguish between a smart perception and a not-so-smart perception by the following example. I open my eyes and I see a certain unrecognized red mass with a specific shape just in front of me. My eyes are working fine, thank you. My visual cortex is processing all of the preliminary visual information, and what vision delivers is the meaningless red mass, which I then have to interpret in some non-visual, non-perceptual cognitive steps that go beyond perception itself. Let us call this a not-so-smart perception. In contrast, in the very same situation, when I open my eyes I see my car. It is true that it has a specific shape and is red, and I do see the shape and the color, but I see the shape and color as being aspects of something that is amazingly recognizable as my car. Actually, if you ask me what I see, I would likely not say that I see a red and shapely mass. Somehow I see through those aspects and I see my car. I do not see red mass, shape, and color, and then try to piece all of that together to make it add up to my car. I simply and directly see my car. So let us call this a perception with some degree of smarts.

[1] On average, children who are less than four years are unable to distinguish between how things really are in the world and what other people may falsely believe about such things. According to the standard view, around four years of age we begin to recognize that other individuals have their own sets of beliefs and intentions that inform their behavior. Typically in false-beliefs tests children are asked to observe rather than interact with others.

One way to characterize this smartness is to say that my perception of my car requires that I have a concept of car, and that concept is precisely what informs my perception. This conceptual aspect is what makes the perception smart. I want to set this particular characterization aside for the present and return to it below.

Second, what precisely do we mean by direct? Is this in contrast to an *indirect* perception—and what precisely is an indirect perception? It turns out that there are not many examples of indirect perception. In one sense, perception of something through a mirror might be considered an indirect perception of that thing. I directly see the mirror image, and that image is of some X. So I am indirectly seeing X. Or in the tactile modality, perhaps I feel the shape of something through something else, e.g., a cloth that loosely shrouds the shape of something. A blind person may have a tactile perception of the environment through the use of a cane, or a scientist might perceive something by means of an instrument. But I will set these examples aside as irrelevant to the current discussion, and in doing that I suggest that what we mean by direct perception (or a direct perceptual grasp) is nothing more than perception itself, not in contrast to indirect perception, but in the sense that all perception is direct. Even when I see something through a mirror I am directly seeing the mirror image; the blind person's cane tends to disappear, experientially, and become part of his body schema much like his natural sense organs. The relevant contrast is not between direct and indirect perception but between perception and something added to perception, e.g., an inference or interpretation that goes beyond what is perceived. If I directly see my car I do not ordinarily have to make an inference on the basis of what I see that it is my car. Of course, there may be a case in which I would have to make such an inference. For example, if my car was terribly totaled in an accident, I may not recognize it at first and I may have to use certain clues about its appearance to infer that it is my car. We will come back to this point because something similar might happen in the case of social perception.

Of course I am not denying that when I see my car all kinds of complex processing is going on in my brain.[2] The visual cortex is processing information about shape and about color, and so forth, in a distributed fashion. Neuroscientifically I may in fact be able to carve up this early processing in the visual system where shape and color are processed in V1 and intermediate visual areas (e.g., Gustavsen & Gallant, 2003). More processing in the inferior temporal cortex, and "top-down" processes may focus and integrate the sensory information. Perhaps the neuroscientist even thinks there is a binding problem and is motivated to ask how all this distributed processing gets glued together to form a coherent perception of my car. But that is a problem for the neuroscientist; not for the perceiver. We want to distinguish between a sub-personal explanation of perception, and conscious perception itself. At the personal or conscious level, I do not have to perceptually piece together the shape and the color and the mass in order to get my car. Even if the sub-personal processes are complex (and I do not deny that they are), the perception that I have of my car is direct—I see it right there in front of me. I do not have to glue anything together, add an interpretation or add an inference.

Indeed, the sub-personal processes involved here are even more complex than any reference to the binding problem involves. If I recognize something *as* something, then clearly some kind of memory and anticipatory processing is involved. This does not mean, for example, that I have to explicitly or consciously remember that I have a car or what it looks like. Memory works implicitly to inform my perception. But this informing process is already built into the perceptual process so that as I consciously perceive, my perception is already informed by the relevant sub-personal processing. I don't first perceive and then add memory in order to recognize my car. My perception, in this sense, is direct even if the sub-personal sensory processing that underpins it follows a complex and dynamic route.[3]

There is more complexity to add when we consider that perception is enactive. I see the car not just as some object among others, but as an object that I can use—that I can climb into and drive. We know that when we perceive an instrument that we can grasp or manipulate canonical neurons in our pre-motor cortex are activated (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). There is likely a more complex activation taking place when I see a useful machine such as my car. It affords certain kinds of action, and these affordances inform—are part and parcel—of my perceptual process. I see my car *as* drivable. This does not mean that I see my car, and then judge that it is drivable. A separate judgment is not required, unless, of course, the situation is unusual. For example, if my car is badly damaged I may have to inspect it and reach some kind of judgment that it is drivable. Ordinarily, however, the drivability of my car is built into my direct perception of it.

[2] The term 'direct perception' is often associated with Gibsonian psychology, and on one version of Gibson's theory all of the information that we need is simply picked up from the environment. Although I favor a non-representational and non-inferential view of perception (contra Fodor as well as Helmholtz), I do not deny that the organism, including the brain, has something to contribute to the shaping of perceptual information. Perception does involve complex processes at a sub-personal or unconscious level, but this does not make perception, which is an activity of the perceiving subject (organism), indirect. For a Gibsonian view that is closer to the view of direct perception that I am defending here, see Tony Chemero's account informed by situation semantics in Chemero (2006).

[3] Does this sub-personal activity involve inference? Inference in some minimal sense involves moving from some known information, which one takes as a clue or as evidence, to something that is not known. First, it is important to note that neurons do not face the world alone. They function only as part of a large and complexly interconnected system. Even a mirror neuron that fires when I see an intentional action (even an incomplete or partially occluded action) is not functioning on its own—it fires only if a variety of other neurons and neuronal systems are working in a certain way. When mirror neuron x fires, it is not just because I see action A, but also because all kinds of neuronal activity is going on, including, of course, activity in the visual cortex, and under conditions defined by a host of other factors, including levels and effects of neurotransmitters. Furthermore, it is misleading to say that the neuron is responding to the action, since the neuron doesn't observe the action. Sub-personal processes do not observe or perceive; and even non-conscious perception is something that the organism as a whole does. On what basis, then, could there be anything like sub-personal inference?

Add to this a certain emotional coloration that may come along with my perception. I may have a love–hate relation with my car. This does not necessarily generate an explicit feeling every time I see my car, but it may certainly inform the way that I perceive it. Studies have shown that the fusiform gyrus "face area" of the brain is activated, not only for face perception, but also when we look at the front (grill, headlights) of cars (Gauthier, Skudlarski, Gore, & Anderson, 2000; Xu, 2005). Whatever the significance of this, would we not say that such neural activation is part of what underpins the perception that I have of my car if in fact I am looking at its front end? Activation of the fusiform gyrus in such cases is not the underpinning of some additional cognitive act. I do not perceive and then entertain thoughts about how closely my car resembles my best friend, or myself. Nor is it justified to say that I sub-personally infer something about a resemblance between my car and something else. That kind of claim is itself an ungrounded inference. Rather, it seems more reasonable to say that activation in fusiform gyrus when viewing cars may reflect or contribute to the implicit emotional coloration of my perception.

Meaningful perception of any sort may rely on activation of association brain areas outside of very early perceptual (e.g., V1 for visual) processing areas. Recent research, however, has shown that even neuronal activity in the earliest of perceptual processing areas, such as V1, reflects more than simple feature detection. V1 neurons anticipate reward if they have been tuned by prior experience (Shuler & Bear, 2006). Once again it is not perception first, followed by a higher cognitive function that registers the possibility of reward.

Perception (of a car, my place of work, my home, a work of art, a landscape, instruments, objects of various kinds, as well as events) is a highly complex phenomenon. It certainly involves complex and articulated processes on the sub-personal level.[4] At a reflective phenomenological level we may be able to analyze our perceptual experience in terms of its enactive and emotional content. Pre-reflectively, however, that is, as we live through our perceptions, our experience is a richly informed direct grasp of whatever is presented.

There is no suggestion here that direct perception is anything like a "pure" perception. Just the opposite. The question about direct perception, whether of objects or of others is not simply about how direct it is, or what directness means, but how smart, how richly informed, it is. The smarter the perception is, the more work it does; the dumber it is, the more it requires extra cognitive processes (theory, simulation) to get the job done. The direct perception theorist is claiming that social perception is very smart and that in the usual circumstances of social interaction it does most of the work without the need of extra cognitive (theoretical or simulationist) processes.

3. Direct perception involves neither theory nor simulation

The concept of direct intersubjective perception that I want to defend is one that involves a relatively smart perception. It is captured by the following statement from Max Scheler.

> For we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands... And with the tenor of this thoughts in the sound of his words. If anyone tells me that this is not 'perception', for it cannot be so, in view of the fact that a perception is simply a 'complex of physical sensations... I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts (Scheler, 1954, pp. 260–261).

> One finds this idea in Wittgenstein as well. Look into someone else's face, and see the consciousness in it, and a particular shade of consciousness. You see on it, in it, joy, indifference, interest, excitement, torpor, and so on... Do you look into yourself in order to recognize the fury in *his* face? (Wittgenstein, 1967, § 229).

> In general I do not surmise fear in him—I *see* it. I do not feel that I am deducing the probable existence of something inside from something outside; rather it is as if the human face were in a way translucent and that I were seeing it not in reflected light but rather in its own. (Wittgenstein, 1980, § 170).

> "We *see* emotion."—As opposed to what?—We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. Grief, one would like to say, is personified in the face. (Wittgenstein, 1980, § 570,; see Overgaard, 2005, for discussion).

How smart is this kind of perception? If we are measuring the IQ (so to speak) of perception, then it is not just a matter of how much memory or previous experience has informed it. Previous experience can tune our sensory-motor neuronal systems; association areas in the brain can integrate memory and emotion with sensory processes. As the empiricists from the time of Locke suggest, perception needs to be educated by experience. At the same time, however, we do not arrive in the world as a *tabula rasa*—and our slate starts to fill up very quickly. Developmental studies consistently tell us that neonate perception is already relatively smart. The newborn infant can pick out a human face from the crowd of objects in its environment, with sufficient detail that will enable it to imitate the gesture it sees on that face (see Gallagher & Meltzoff, 1996;

[4] Much of what I have said is focused on brain processes. But it also the case that the body as a whole, and the pre-processing that goes on in the peripheral nervous system, plays an important role in shaping and informing perception. See e.g., Chiel & Beer 1997; Gallagher 2005, Chap. 6.

Meltzoff & Moore, 1977, 1994). There is increasing evidence that we automatically attune to smiles (and other facial gestures) with an enactive, mimetic, response (Schilbach, Eickhoff, Mojzisch, & Vogeley, 2008).[5] The young infant is visually attracted to movement and in specific ways to biological movement, and auditorily attracted to certain kinds of sounds, such as its mother's voice. Infants "vocalize and gesture in a way that seems [affectively and temporally] 'tuned' to the vocalizations and gestures of the other person" (Gopnik & Meltzoff, 1997, p. 131). Not only do human infants show a wide range of facial expressions, complex emotional, gestural, prosodic, and tactile face-to-face interaction patterns, absent or rare in non-human primates (Falk, 2004; Herrmann, Call, Hare, & Tomasello, 2007), but notably without the intervention of theory or simulation, and in a non-mentalizing way, they are able to see bodily movement as expressive of emotion, and as goal-directed intentional movement, and to perceive other persons as agents. This does not require advanced cognitive abilities, inference, or simulation skills; rather, it is a perceptual capacity that is 'fast, automatic, irresistible and highly stimulus-driven' (Scholl & Tremoulet, 2000, p. 299).

At this point we need to come back to the issue of whether perception, as we are characterizing it, is conceptual, since there is good reason to think that infants this young do not yet have concepts (see, e.g., Bermúdez, 1998).[6] How we answer this depends to some extent on an account of concept acquisition. Here, I will assume an account that links concept acquisition to language acquisition. Perception in early infancy starts out, accordingly, as non-conceptual. This does not mean that in early perceptual experience we do not find distinctions that correspond to distinct entities encountered in the world. Faces are different from other things; agents are different from non-agents; my mother's voice is different from other voices—and these differences are experienced despite the fact that the infant has no concept of face, agent, or voice. As Scheler puts it, the environment is experienced as immediately parsed into entities that manifest expressiveness and others that do not (Scheler, 1954). Non-conceptual perception is still relatively smart; and it gains in intelligence as the infant develops, acquires language, conceptual competency, and narrative competency. There is no doubt that advances associated with language and concept acquisition will transform perceptual experience, and specifically along lines that are pragmatic and intersubjective, some of which are already traced out in early non-conceptual experience—faces remain faces; agents remain agents; voices remain voices. As a result, by the time I have reached my 18th birthday (and likely long before that) I have no problem seeing, in a very direct way, that this particular agent with this particular face and this particular voice is driving a very desirable car. Yet, neither the smartness nor the directness of perception is necessarily dependent on perceiving things under concepts.

Infant social perception, then, can be non-conceptually smart. Very smart. Infants are able to detect correspondences between visual and auditory information that specify the expression of emotions as early as 5–7 months (Walker, 1982; also, Hobson, 1993, 2002). At 9 months infants follow the other person's eyes (Senju, Johnson, & Csibra, 2006), and start to perceive various movements of the head, the mouth, the hands, and more general body movements as meaningful, goal-directed movements. Baldwin and colleagues, for example, have shown that infants at 10–11 months are able to parse some kinds of continuous action according to intentional boundaries (Baird & Baldwin, 2001; Baldwin & Baird, 2001; Baldwin, Baird, Saylor, & Clark, 2001). Such perceptions give the infant, by the end of the first year of life, a non-mentalistic, perceptually based embodied understanding of the intentions and dispositions of other persons (Allison, Puce, & McCarthy, 2000; Baldwin, 1993; Johnson, 2000; Johnson & et al., 1998). And at 18 months infants can re-enact to completion the goal-directed behavior that an observed subject does not complete, showing that they recognize the unfulfilled intentions of others (Herrmann et al., 2007; Meltzoff, 1995).[7] These capabilities do not disappear in adulthood; they mature and become more sophisticated (see e.g., Dittrich, Troscianko, Lea, & Morgan, 1996).

If the question is how far perception can go, it seems that it can reach to intentions and feelings. Many theorists will claim, however, that intentions and feelings are not things that can be seen. They are mental states. Perception cannot be smart enough to reach the mental states that are hidden away (imperceptible) in the other person's mind.

The best way to address this question is to come at it from the other side. What is it that we do grasp when we understand the other person. Do we really mindread, in the sense of composing something like an inference from the observed behavior

[5] The sub-personal complexity of seeing another person's smiling face is captured by Schilbach et al. When we see a smiling face we automatically, involuntarily, and non-consciously mimic the smile, at least in terms of specific muscle activation. "The neural network active during this period... i.e., during the occurrence of facial mimicry, is formed by significant activations in the precentral motor area, as well as in several other, non-motor regions of the brain. More precisely, our analysis demonstrated activations pertaining to the window of interest within the posterior portions of cingulate gyrus on both hemispheres. Furthermore, differential neural activity is seen in the right hippocampus. Differential neural activity was also observed in the dorsal midbrain,... interpreted as activity in the midbrain tectum. Additionally, our results demonstrate activation of the inferior aspects of the left precentral gyrus, i.e., the face motor area. Lastly, differential brain activity is found in the left precuneus..." (2008, 42). Schilbach et al. suggest that "the process of perceiving faces always includes an 'enactive' element through which we engage with and respond to stimuli instead of a mere 'passive' perception of face-based cues" (2008, 45).

[6] My thanks to one of my referees for bringing this question to my attention.

[7] I cite these studies for the data presented rather than for the authors' interpretation of that data which tends to be fully in line with ToM approaches. The data, however, suggests that the infants actually see that the other person has intentions that are not completed and recognizes them as such. The study by Meltzoff shows this at 18 months; the study by Herrmann et al. shows this in 2.5 year olds. Onishi and Baillargeon (2005) have recently shown that infants at 15 months apparently mentalize the false beliefs of others. The data from their experiments suggest that infants see what the other person intends to do and is surprised (or at least notices) when the behavior of the other violates what the infant knows about the context (specifically about who has seen or not seen certain events). Although Onishi and Baillargeon interpret the data entirely within a ToM framework of mentalizing the other's beliefs, an alternative interpretation in terms of perceived meaningful (contextualized) behavior, actions, and intentions is clearly available. See, e.g., Woodward, & Sommerville (2000): "[...] 12-month-old infants interpreted action in context in two senses: They used both the other actions performed by the actor and the causal constraints in the situation to interpret an ambiguous action... infants as young as 6 months construe grasping as goal directed, infants under 12 months may be able to interpret the goal of an action on the basis of sequences [of actions in context]" (p. 76–77). Appeal to hidden beliefs or mental states is not required. Also see Király, Jovanovic, Prinz, Aschersleben, and Gergely (2003) and Biro, Csibra, and Gergely (2007).

to a set of inaccessible mental states? This certainly seems to be how TT and ST picture it. This is how they define the problem to be solved. How do we move from perceived behavior to the hidden mental states?[8] But that question is itself put into question by the direct perception approach. This approach suggests that in our normal everyday encounters, this is not what we do. We do not ordinarily need to go further than what is already the rich and complex comprehension that we gain through the perception of a situated agent—that is, of an agent who is situated in an environment which also tells us something about what that person is doing and thinking. If I see the situation and what the agent is doing in it, and how the agent is doing it, and what the agent is expressing (e.g., through her gestures and style of movement), and this perception is already informed by my own interaction with them and others, as well as by my previous situated experiences, my habitual ways of understanding, and by cultural norms and established practices, and so forth, then in cases which we encounter in our normal ordinary engagements the work of understanding is already sufficiently accomplished and I do not have to go any further. I do not have to start thinking about what might be going on in the other person's mind since everything I need for understanding her is there in her action and in our shared world.[9]

Indeed, in ordinary instances of interaction with others, I am not in the observer position; I am not off to the side thinking or trying to figure out what they are doing. Rather, I am responding to them in an embodied way. What we call social cognition is often nothing more than that social interaction. What I perceive in these cases does not constitute something short of understanding. Rather my understanding of the other person is constituted within the perception–action loops that define the various things that I am doing with or in response to others. This can be clearly shown in a micro-analysis of the postures, movements, gestures, gazes, and facial expressions of people as they engage in a novel task and where communication among them comes in the very actions that they take (see Lindblom, 2007; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005).

Still, not all behavior can be perceptually grasped; behavior is often ambiguous. Ambiguity is reduced, however, and our sense of the other person's behavior is helped along by context—which may be physical, social, cultural, or all of these—and by interaction. This does not rule out the possibility of unresolved ambiguity, or that the other person may in some circumstances be a real puzzle. I may not have enough information perceptually to make sense out of what the other person is doing. Other persons do not always reveal their thoughts or emotions via their expressions, actions, behaviors, etc. The claim here is not that direct perception (or other aspects of this approach—the reliance on context, social roles, narrative, etc.) can penetrate to the soul of the other person and discover her innermost emotional states. Nor is the claim that we can never be misled by what we perceive. The claim is rather that for the most part, in most of our encounters in everyday life, direct perception delivers sufficient information for understanding others. In addition it would only be something that I discover via these means that would lead me to the idea that perhaps something more is going on with the other person. At that point I would not hesitate to say that we might then turn to other means (narrative or even theory or simulation) to more deeply interpret the person. The point is, however, we first have to recognize the person's behaviors or expressions as perplexing or mysterious, and it would be this that would motivate further interpretation. Furthermore, I claim that this is relatively rare. I don't constantly go around trying to solve such puzzles (unless, for example, I am high-functioning autistic—see Gallagher, 2004). Practically speaking, direct perception, etc. delivers what I need to interact with others most of the time. In the broad range of normal circumstances there is already so much available in the person's movements, gestures, facial expressions, and so on, as well as in the pragmatic or social context, that I can grasp everything I need for understanding in what is perceptually available (see Ratcliffe, 2007).

4. Direct perception or implicit simulation

Perception is direct and smart. But it is also innately (or very early) tuned to socially relevant aspects of the world. In contrast to this concept of direct social perception, on the TT and standard ST accounts it seems that perception is only one part of a larger cognition process and that the social aspect is really found in what follows on from perceptual observation, in the application of folk psychology or in simulation routines, but it is not there in perception itself. This is where the more recent work on implicit ST and mirror neurons may offer some insight.

Implicit ST appeals to recently developed neuroscientific evidence involving sub-personal activation of mirror neurons, shared representations, or more generally, resonance systems. My motor system is activated when I perceive another person performing an intentional action, for example. Mirror neurons in the pre-motor cortex, in Broca's area, and in the parietal cortex of the human brain are activated both when the subject engages in specific instrumental actions, and when the subject observes someone else engage in those actions (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996; Rizzolatti, Fogassi, & Gallese, 2000). Also, specific overlapping neural areas (shared representations), in parts of the frontal and parietal cortexes, are acti-

[8] Some versions of this view do refer to interaction, but stay with the idea that what is at stake is the interaction of two minds that have to communicate across the seemingly thin air of an unbridgeable gap. For example, "... the study of social interaction... is concerned with the question of how two minds shape each other mutually through reciprocal interactions. To understand interactive minds we have to understand how thoughts, feelings, intentions, and beliefs can be transmitted from one mind to the other" (Singer, Wolpert, & Frith 2004, p. xvii). Contrast this to the view expressed by Gibbs (2001, p. 120): "[T]he intentionality in the mother-infant interaction does not reside in any individual mind; it emerges as a product of their social interaction. Thus, what is intentional about the mother–infant interaction cannot be explained simply in terms of the mother's and infant's intentions with respect to each other."

[9] This does not rule out extra-perceptual inference for more complex actions that may involve relatively long-term or overarching goals. See, e.g., Woodward and Sommerville (2000).

vated when I engage in intentional actions and when I observe some other person engage in that action (e.g., Grèzes & Decety, 2001). How should we interpret this evidence? Some simulation theorists might claim that these processes underpin (or are the neural correlates) of explicit acts of simulation (see e.g., Jeannerod & Pacherie 2004, p. 129). *Implicit* simulation theorists, however, contend that these sub-personal processes themselves just are a simulation of the other's intentions. On this view, activation of mirror neurons involves "automatic, implicit, and non-reflexive simulation mechanisms..." (Gallese, 2005, p. 117; also see Gallese, 2007). According to Gallese's "shared manifold hypothesis" our empathic experience of the other person at the phenomenological level is underpinned by the activity of "mirror matching neural circuits" at brain level, and these are interpreted as "simulation routines, *as if* processes enabling models of others to be created" at the functional level (2001, p. 45). On this hypothesis, at the explicit, phenomenological level, one is not explicitly (consciously) simulating; rather the simulation process remains entirely at the sub-personal level.

Implicit ST understood in these or in similar terms is fast becoming the consensus view. A growing number of researchers take up the simulation terminology as a generally accepted way to describe mirror system processes. For example, Jean Decety and Grèzes (2006, p. 6), explaining Rizzolatti's position, put it this way:

> By automatically matching the agent's observed action onto its own motor repertoire without executing it, the firing of mirror neurons in the observer brain simulates the agent's observed action and thereby contributes to the understanding of the perceived action.

Goldman (2006) now distinguishes between simulation as a high-level (explicit) mind-reading and simulation as a low-level (implicit) mind-reading where the latter is "simple, primitive, automatic, and largely below the level of consciousness" (p. 113), and the prototype for which is "the mirroring type of simulation process" (p. 147). That mirror neuron activation is a simulation not only of the goal of the observed action but of the intention of the acting individual, and therefore a form of mind-reading, is suggested by research that shows mirror neurons discriminate identical movements according to the intentional action and contexts in which these movements are embedded (Fogassi et al., 2005; Iacoboni et al., 2005; Kaplan & Iacoboni, 2006). Neural simulation has also been extended as an explanation of how we grasp emotions and pain in others (Avenanti & Aglioti, 2006; Gallese, Eagle, & Migone, 2007; Minio-Paluello, Avenanti, & Aglioti, 2006). The idea that "simulator neurons" are responsible for understanding actions, thoughts, and emotions is taken up by Oberman and Ramachandran (2007) who amass evidence that the mirror-neuron system as an internal simulation mechanism is dysfunctional in cases of autism.

I have suggested elsewhere (Gallagher, 2007a; Gallagher, 2007b), however, that there are conceptual problems involved in calling sub-personal mirror resonance processes 'simulations'. These sub-personal processes fail to meet the definition of simulation as it is developed in ST. On that definition, simulation involves a kind of instrumental control as we use a model to understand something that we cannot understand directly. In addition, simulation involves pretense—the idea that we use our own mental states "as if" they were the mental states of others. In contrast, the sub-personal mirroring processes, (1) rather than having an instrumental character under our control, are automatic and elicited by the actions of others. Moreover, (2) because mirror neurons are activated both when I act and when I see someone else act, they are neutral with respect to who the agent is (deVignemont, 2004; Gallese, 2005; Hurley, 2005; Jeannerod & Pacherie, 2004). As a result, sub-personal processes do not and can not involve pretense, which requires distinguishing one agent (me) from another (you). A weakened or minimal definition of simulation as simply a form of matching which jettisons the instrumental and pretense aspects (Goldman, 2006; Goldman & Sripada, 2005) fails to explain how we understand others who are engaged in very different activities from us, or who are experiencing very different emotions. In addition, there is neuroscientific evidence that shows that mirror neuron activation does not necessarily involve a precise match between motor system execution and observed action, but may be involved in "logically related" actions or in anticipating future action (Csibra, 2005; Iacoboni et al., 2005).

To deny that mirror resonance processes constitute simulations is not to deny that mirror neurons may be involved in our interactions with others, possibly contributing to our ability to understand others or to keep track of ongoing intersubjective relations. Rather, directly relevant to the concept of direct perception, I want to suggest that an alternative and more parsimonious interpretation of mirror neuron activation is possible. Mirror resonance processes can easily be interpreted as part of the neuronal processes that underlie social *perception*. That is, the articulated neuronal processes that include activation of mirror neurons or shared representations constitute the neural correlates of a non-articulated immediate perception of the other person's intentional actions, rather than a distinct process of simulating their intentions.[10]

Perception of an unfolding gesture or movement does not happen in an instant; it is a temporal phenomenon. Mirror neurons fire 30–100 ms after appropriate visual stimulation. This short amount of time between activation of the visual cortex and activation of the mirror-neuron system raises the question of where precisely to draw the line between perceptual processes and something that would count as something more than the perceptual processes. A distinction at the neural level between activation of the visual cortex and activation of the pre-motor cortex does not mean that this constitutes a distinction between processes that are purely perceptual and processes that involve something more than perception.

Furthermore, if we think of perception as an enactive process (Hurley, 1998; Noë, 2004; Varela et al., 1991), as involving sensory–motor skills rather than as just sensory input/processing—as an active, skillful, embodied engagement with the

[10] Since mirror neurons are found in monkeys, does this mean that they are capable of social perception? Given the status of current debate on this kind of issue, I will leave this as an open question. The important thing, however, is to note that mirror neurons are only part of the story and likely not sufficient for social perception of intentions. My thanks to one of my referees for calling this issue to my attention.

world rather than as the passive reception of information from the environment—then it is certainly more appropriate to think of mirror resonance processes as part of the structure of the perceptual process when it is a perception of another person's actions. Mirror activation is not the initiation of simulation; it subtends a direct intersubjective perception of what the other is doing.

This interpretation of mirror neuron activation provides a tight fit with the direct perception account of intersubjective understanding and interaction. These mirroring processes underpin a capacity for human interaction and intersubjective understanding, already operative in infancy in certain embodied practices—practices that are emotional, sensory–motor, non-conceptual, and directly perceptual—practices that involve a perceptual sense of others and that can come to constitute a common bodily intentionality shared by both the perceiving subject and the perceived other (Gallagher, 2001, 2005). On this view the mind of the other person is not something that is entirely hidden away and inaccessible. In seeing the actions and expressive movements of the other person in the context of the surrounding world, one already sees their meaning; no inference to a hidden set of mental states (beliefs, desires, etc.) is necessary. When I see the other's action or gesture, I see (I *immediately perceive*) the meaning in the action or gesture. I see the joy or I see the anger, or I see the intention in the face or in the posture or in the gesture or action of the other.

This alternative, non-simulationist interpretation of the neuroscience of mirror neurons, then, suggests that before we are in a position to theorize, simulate, explain or predict mental states in others, we are already in a position to interact with and to understand others in terms of their expressions, gestures, and purposive movements, reflecting their intentions and emotions. We already have specific perception-based understandings about what others feel, whether they are attending to us or not, how they are acting toward us and others, whether their intentions are friendly or not, and so forth; and in most cases we have this without the need for personal-level theorizing or simulating about what the other person believes or desires. Moreover, we have this without the benefit of anything that on the sub-personal level could be considered an extra cognitive step, a simulation, or inference.

My goal has not been to give a full account of social cognition, but to suggest that one important element of our understanding of others depends on a direct perception of the other person's actions and expressive movements. In this regard direct perception is nothing mysterious. It is an enactive and intersubjectively attuned perception that on the sub-personal, neural level depends at least in part on mirror resonance processing. For a fuller account of social cognition and interaction in its more sophisticated and nuanced forms, one also must include the roles played by emotion, context, and culture; contributories that are frequently shaped and delivered by narrative processes (see Gallagher & Hutto, 2008).

References

Allison, T., Puce, Q., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science, 4*(7), 267–278.

Avenanti, A., & Aglioti, S. M. (2006). The sensorimotor side of empathy for pain. In M. Mancia (Ed.), *Psychoanalysis and neuroscience* (pp. 235–256). Milan: Springer.

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 193–206). Cambridge, MA: MIT Press.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Science, 5*(4), 171–178.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717.

Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, MA: MIT Press.

Biro, S., Csibra, G., & Gergely, G. (2007). The role of behavioral cues in understanding goal-directed actions in infancy. *Progress in Brain Research, 164*, 303–322.

Chemero, A. (2006). Information and direct perception: A new approach. In P. Farias & J. Queiroz (Eds.), *Advanced issues in cognitive science and semiotics* (pp. 59–72). Aachen: Shaker Verlag.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences, 20*, 553–557.

Csibra, G. 2005. Mirror neurons and action observation. Is simulation involved? ESF Interdisciplines. Available from <http://www.interdisciplines.org/mirror/papers/>.

Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research, 1079*, 4–14.

deVignemont, F. (2004). The co-consciousness hypothesis. *Phenomenology and the Cognitive Sciences, 3*(1), 97–114.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*, 727–738.

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences, 27*(4), 491–503.

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

Gallagher, S. (2007a). Logical and phenomenological arguments against simulation theory. In D. Hutto & M. Ratcliffe (Eds.), *Folk psychology re-assessed* (pp. 63–78). Springer Publishers.

Gallagher, S. (2007b). Simulation trouble. *Social Neuroscience, 2*(3-4), 353–365.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

Gallagher, S. (2001). The practice of mind: Theory, simulation, or interaction? *Journal of Consciousness Studies, 8*(5–7), 83–107.

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, and Psychology, 11*(3), 199–217.

Gallagher, S., & Hutto, D. (2008). Primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins.

Gallagher, S., & Meltzoff, A. N. (1996). The earliest sense of self and others. *Philosophical Psychology, 9*, 213–236.

Gallese (2005). 'Being like me': Self-other identity, mirror neurons and empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation I* (pp. 101–118). Cambridge, MA: MIT Press.

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

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

Gallese, V., Eagle, M. N., & Migone, P. (2007). Intentional attunement: Mirror neurons and the neural underpinnings of interpersonal relations. *Journal of the American Psychoanalytic Association, 55*(1), 131–176.
Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197.
Gibbs, R. W. (2001). Intentions as emergent products of social interactions. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 105–122). Cambridge, MA: MIT Press.
Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. Oxford, England: Oxford University Press.
Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*(2005), 193–213.
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1–19.
Gustavsen, K., & Gallant, J. L. (2003). Shape perception: complex contour representation in visual area V4. *Current Biology, 13*(6), R234–R235.
Herrmann, E., Call, J., Hare, B., & Tomasello, M. (2007). Humans evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366.
Hobson, P. (1993). The emotional origins of social understanding. *Philosophical Psychology, 6*, 227–249.
Hobson, P. (2002). *The cradle of thought*. London: Macmillan.
Hurley, S. L. (2005). Active perception and perceiving action: The shared circuits model. In T. Gendler & J. Hawthorne (Eds.), *Perceptual experience*. New York: Oxford University Press.
Hurley, S. L. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.
Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Giacomo Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3*(3), 529–535.
Jeannerod, M., & Pacherie, E. (2004). Agency, simulation, and self-identification. *Mind and Language, 19*(2), 113–146.
Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Science, 4*, 22–28.
Johnson, S. et al (1998). Whose gaze will infants follow? The elicitation of gaze-following in 12-month-old infants. *Developmental Science, 1*, 233–238.
Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. *Social Neuroscience, 1*(3–4), 175–183.
Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition, 12*(4), 752–769.
Lindblom, J. 2007. *Minding the body: Interacting socially through embodied action*. Linköping: Linköping Studies in Science and Technology, Dissertation No. 1112.
Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.
Meltzoff, A., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.
Meltzoff, A., & Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development, 17*, 83–99.
Minio-Paluello, I., Avenanti, A., & Aglioti, S. M. (2006). *Social Neuroscience, 1*(3–4), 320–333.
Niedenthal, P. M., Barsalou, L. M., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184–211.
Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.
Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin, 133*(2), 310–327.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258.
Overgaard, S. (2005). The problem of other minds: Wittgenstein's phenomenological perspective. *Phenomenology and the Cognitive Sciences, 5*(1), 53–73.
Ratcliffe, M. J. (2007). *Rethinking commonsense psychology: A critique of folk psychology. Theory of Mind and Simulation*. Basingstoke: Palgrave Macmillan.
Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 539–552). Cambridge, MA: MIT Press.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.
Scheler, M. 1954. *The nature of sympathy*. London: Routledge. Kegan Paul, 1923. Original: *Wesen und Formen der Sympathie*. Bonn: Verlag Friedrich Cohen.
Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience, 3*(1), 37–50.
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 299–309.
Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience, 1*(3–4), 220–234.
Shuler, M. G., & Bear, M. F. (2006). Neurons in the primary visual cortex respond differently to a flash of light after it has been paired with a reward, unexpectedly showing that cognitive information is coded at this level in the cortex. *Science, 311*, 1606–1609.
Singer, W., Wolpert, D., & Frith, C. (2004). Introduction: The study of social interactions. In C. Frith & D. Wolpert (Eds.), *The neuroscience of social interaction* (pp. xii–xxvii). Oxford: Oxford University Press.
Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge: MIT Press.
Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33*, 514–535.
Wittgenstein, L. 1980. *Remarks on the philosophy of psychology*, In G. H. von Wright & H. Nyman (Eds.) (Vol. II) (C. G. Luckhardt & M. A. E. Aue, Trans). Oxford: Blackwell.
Wittgenstein, L. 1967. *Zettel*. In G. E. M. Anscombe & G. H. von Wright (Eds.) (G. E. M. Anscombe, Trans.). Berkeley: University of California Press.
Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science, 11*, 73–77.
Xu, Y. (2005). Revisiting the role of the fusiform and occipital face areas in visual expertise. *Cerebral Cortex, 15*, 1234–1242.