La perception d'autrui en action. *Fondements cognitifs de l'interaction avec autrui*. Collège de France (22 February 2006).

Perceiving others in action / La perception d'autrui en action[1]

Shaun Gallagher

In a *New York Times* article last month, entitled "Cells that read minds," the neuroscience reporter, Sandra Blakeslee (January 10, 2006) provided a list of all the things that mirror neurons can explain. As we know, mirror neurons, discovered by Rizzolatti's group in Parma, are neurons that are activated when we engage in action, and when we perceive intentional movement in another person. According to Blakeslee and the scientists she interviewed, mirror neurons explain not only how we are capable of understanding another person's actions, but also language, empathy, "how children learn, why people respond to certain types of sports, dance, music and art, why watching media violence may be harmful and why many men like pornography." Let me set aside the controversial questions about whether mirror neurons can explain all of these things, and accept that mirror neurons are clearly smart little cells. But let me ask whether Blakeslee and her scientists are expressing things in the right way.

My worry, unlike some philosophers, is not so much about the language per se, or about how they phrase what they say, although I think this is clearly a problem. Rather I'm worried about the phenomenology, and about a categorical confusion between phenomenology and neurology. This confusion is obvious in the way Blakeslee puts it: mirror neurons can *analyze* scenes and *read minds*, they *understand* actions and intentions. Perhaps this is excusable as a kind of journalistic license or shorthand for the idea that mirror neurons allow *the person* to analyze scenes, read minds, and understand actions and intentions. Scientists are sometimes more careful, but, as I will argue, sometimes their careful language hides a similar kind of confusion. Rizzolatti, for example, tells Blakeslee that "Mirror neurons allow us to grasp the minds of others not through conceptual reasoning but through direct simulation. By feeling, not by thinking." And Marco Iacoboni is equally as careful: "When you see me perform an action - such as picking up a baseball - you automatically simulate the action in your own brain." Both of these quotations point to the close association that has grown up between the neuroscience of resonance systems (or mirror systems) and the simulation theory. This

[1]This is the text of a lecture that is part of a series entitled *Fondements cognitifs de l'interaction avec autrui* given at the Collège de France, Amphithéâtre Marguerite de Navarre, 11, place Marcelin Berthelot, Paris (February 2006), en anglais avec traduction en français. The lecture series was organized by Professor Alain Berthoz, Laboratoire de Physiologie de la Perception et de l'Action, Collège de France.

association is what I want to examine in some detail, and I want to put it into question. Let me start, however, with an example of the more obvious kind of confusion, which will lead us directly to the question about the relation between resonance and simulation.

Bare naked intentions?

I start with an example of what I think is an obvious confusion between two different levels of explanation – the neurological and the phenomenological. I must confess surprise to find this confusion in two authors, a neuroscientist and a philosopher, whom I consider to be very good thinkers, Marc Jeannerod and Elizabeth Pacherie, In what is otherwise an excellent review article about recent work in the neuroscience of social cognition, these two co-authors make some odd claims about naked intentions.

First, it is important to note that there is a growing consensus that mirror neurons, and more generally, shared representations in those brain areas that are activated when I perform an action and when I perceive another person perform the action, are neutral in regard to determining agency -- they are activated both for my own action and for observation of the other's action: activation of the system registers the intentional action but not the agent (deVignemont 2004; Gallese 2005; Hurley 2005; Jeannerod and Pacherie 2004). The idea is that if the very same neuronal systems in my brain are activated both when *my* body is in action and when I perceive *your* body to be in action, then there must be some other mechanism that allows me to distinguish whether the action in question is mine or yours. One way to solve this problem can be found in the proposal made by Georgieff and Jeannerod (1998) which they term the "Who system." If there is an overlap of activation in the shared representations that registers intentional action, there must be, in addition, some mechanism in the non-overlapping areas that signals that the action is mine rather than yours, or vice versa. This mechanism answers the question: "Whose action is it?" Recent neuroimaging experiments have confirmed this idea, showing differential activation in the *posterior insula* when I am the agent, and in the *right inferior parietal cortex* when the agent is someone else (Farrer et al. 2003; Farrer and Frith 2002; Ruby and Decety 2001). So let's grant that there is some good neuroscience to support the hypothesis of the Who system. My intention here is not to question that neuroscience.

Jeannerod and Pacherie are in agreement that there are two things to be explained: awareness of intentional action on the one hand, and agency on the other. Distinguishing these two things, they speak of "naked intentions" – that is, intentions or intentional actions for which agency is yet to be determined. It is just here, however, that their proposal becomes problematic. They assume that an articulation at the level of neural activations, specifically between those activations responsible for (1) registering the "naked" intention of an action, and (2) registering the agency for the action, means that there is an articulation *in experience* between the perception of intention and the experience of agency. "We can be aware of an intention, without by the same token being aware of whose intention it is. … something more than the sole awareness of a naked intention is needed to determine its author." If in fact the brain can process information about intentions without assigning agency to the intentions, is it legitimate to say that our experience is similarly articulated? Jeannerod and Pacherie suggest that it is.

> When the naked intention one is aware of yields an overt action, the extra information needed to establish authorship may be found in the outside world. The question 'Is this intention mine?' would then be answered by answering the question: 'Is this my body performing the corresponding action?' (140).[2]

Phenomenologically (experientially), however, intentions in almost all cases come already clothed in agency. The "who" question, which is rightly posed at the neurological level, hardly ever comes up at the level of experience, because the neural systems have already decided the issue -- one way or the other. That is, even if I'm wrong about who the agent of an action is (something that may happen in schizophrenic symptoms of delusions of control), I am still experiencing or perceiving the intention as already determined in respect to agency. The wonderful thing about the "Who system" is that it is completely neurological and sub-personal -- and the results of its activation are hardly ever experientially manifested as "making a decision about who did the action." Rather, the results of its activation are experientially manifested as "X's action" where X is either you or me. Indeed, our direct perception is highly reliable in regard to discriminating between self and non-self. Pathologies and oddly arranged experiments may reveal "who" problems, but in normal ecological behavior it is generally clear whose intention/action it is. As we know from philosophers like Wittgenstein, Shoemaker, and Evans, the self-identification question -- "Someone is intending to pick up the apple, is it me?" -- just doesn't come up (in that specific form, i.e, "as subject").

In effect, it is important to realize that there is no necessary isomorphism between the phenomenological level and the neuronal level. So if the neuronal processes can be defined as involving a step-wise process, this does not mean that a step-wise process needs to show up in phenomenology.[3] I'll refer to the kind of confusions we get when we assume that articulations on the sub-personal level must manifest themselves as articulations on the phenomenological level as the *fallacy of supposed isomorphism*.

It is this idea, that there can be articulatory differences at the different levels of explanation, that leads us back to the question of simulation. Jeannerod and Pacherie are also friends of simulation. "We argue in favor of a simulation hypothesis that claims that actions, whether overt or covert, are centrally simulated by the neural network, and that

---

[2] This extremely articulated process, diagrammed in the form of answering separate questions, is also set out in an explanation that resembles Brentano's theory of intentionality and self-awareness. Our perception of a tree involves primary and secondary objects. The primary perceptual object is the tree; the secondary object is myself as perceiver. "We claim that it is like this with the perception of intention: when Mary watches John open the door, she is primarily aware of an intention to open the door, rather than being primarily aware that John intends to open the door. Similarly, when Mary herself intends to open the door, she is primarily aware of an intention to open the door, rather than being primarily aware that she herself intends to open the door. Let us call this awareness of an unattributed or 'naked'intention" (Jeannerod and Pacherie 2004, 116). For a critique of Brentano's theory of intentionality, see Gallagher and Zahavi, forthcoming; Zahavi (1998; 2004). But even if Brentano were right about self-awareness, the idea that we perceive naked intentions in another's action, and secondarily attribute agency, seems here to be based on a supposed isomorphism between subpersonal and personal-phenomenological levels.

[3] On the question about isomorphism between subpersonal and personal levels, see Gallagher 1997; Hurley 2005; Varela 1996.

this simulation provides the basis for action recognition and attribution" (p. 113). They make their view of simulation more precise:

> As far as the understanding of action is concerned, we regard simulation as the default procedure …. We also believe that simulation is the root form of interpersonal mentalization and that it is best conceived as a hybrid of explicit and implicit processes, with subpersonal neural simulation serving as a basis for explicit mental simulation. (p. 129).

Varieties of simulation theory

As Jeannerod and Pacherie note, there are several different versions of simulation theory (ST). I have elsewhere developed arguments against all of these versions (Gallagher, in press -a). For our purposes here, let me focus on the distinction between explicit (personal-level) and implicit (sub-personal) level simulation. I will point to problems with both versions of ST, and also argue that a combination or hybrid of these versions will not avoid the problems.

The different versions of ST are based on different conceptions of what constitutes the simulation involved. In the first version, the simulation is something *explicit*, that is, an exercise of conscious imagination and deliberative inference. This was the original version of ST outlined by Gordon (1986) and Heal (1986) in their separate papers.[4] Alvin Goldman is a very good representative of this position. The *implicit* version of simulation understands it to be a subpersonal process, either a functional mechanism that is cashed out in terms of neuronal processes (e.g., Hurley 2005), or these neuronal processes themselves (e.g., Gallese 2001). Hybrid theories, like the one proposed by Jeannerod and Pacherie, combine, within ST, explicit and implicit processes, and perhaps the most famous of these hybrid theories is put forward by Gallese and Goldman (1998).[5]

In the explicit version of ST, simulation involves conscious or introspective mental states in which I imagine myself in the other's situation and use the model (simulation) that is generated to predict the other's mental states. Explicit ST is unabashedly mentalistic. That is, in attempting to understand the other person's action, the task is taken to be to understand the mental states, beliefs, desires, etc., that define that person's intentions. Goldman, for example, argues that simulation is explicit insofar as it involves a conscious introspective use of the imagination to conceptually manipulate propositional attitudes (beliefs, desires). "When a mindreader tries to predict or retrodict someone else's mental state by simulation, she uses pretense or imagination to put herself in the target's 'shoes' and generate the target state." (Goldman 2005a). According to Goldman, simulation involves three steps.

First, the attributor creates in herself pretend states intended to match those of the

[4] On another version, the simulation is also explicit, but does not involve a deliberative inference. This is Gordon's notion of radical simulation.

[5] I do not mean to confuse these with what are also called hybrid theories that combine some version of ST with TT (see Currie and Ravenscroft 2002; Nichols and Stich 2003; Mitchell 2005; and Saxe 2005)

> target. In other words, the attributor attempts to put herself in the target's 'mental shoes'. The second step is to feed these initial pretend states [e.g., beliefs] into some mechanism of the attributor's own psychology … and allow that mechanism to operate on the pretend states so as to generate one or more new states [e.g., decisions] Third, the attributor assigns the output state to the target …" [e.g., we infer or project the decision to the other's mind]. (Goldman 2005b, 80-81.)[6]

There are both logical and phenomenological arguments that have been made against this explicit version of ST. First, we should note, as do Gordon and Cruz (2003), that explicit ST is similar to an older theory called the argument by analogy. In the first half of the 20th century arguments were raised against this theory by a number of philosophers, including phenomenologists like Max Scheler and analytic philosophers of mind, like Gilbert Ryle. Some of the same arguments work against ST. Ryle, for example, argued that the logic of simulation isn't correct because the idea of imputing to a variety of others what is true of my simulated action ignores the diversity of their actions**.** "[T]he observed appearances and actions of people differ very markedly, so the imputation to them of inner processes closely matching [one's own or] one another would be actually contrary to the evidence" (Ryle 1949: 54). A similar objection to the logic of simulation was raised by Scheler (1954). If I project the results of my own simulation on to the other, I understand only myself in that other's situation, but I don't understand the other.[7] Given the diversity of motives, beliefs, desires, and behaviors in the world, it is not clear why such a simulation process would be at all reliable. Scheler also suggests another argument against ST. The explicit simulation process seems cognitively too complicated. Infants, and perhaps even animals, seem capable of understanding the intentions of others, but it would be difficult to attribute the complex cognitive processes involved in simulation to them.[8]

Let me propose one more argument against explicit ST. I will call this the simple

---

[6] One might think that there would be problem with the very first step: "the attributor creates in herself pretend states intended to match those of the target." This suggests that the simulator already has some idea of what's going on with the other person. The question then is where does that knowledge come from and why isn't that already the very thing we are trying to explain. Hybrid theorists who combine TT and ST suggest that folk psychology provides, not a sense of what is going on with the other person, but some general rules about how people think and behave in a certain situations, and that this is what the simulationist can use to generate the pretend mental states needed for the simulation process (e.g., Currie and Ravenscroft 2003). In contrast, Goldman appeals to sub-personal mirror resonance processes (discussed below), although he then faces the problem of how to translate these processes into a conceptual grasp of propositional attitudes. He proposes to solve the latter problem by an appeal to phenomenal properties of propositional attitudes (2002). Goldman suggests that a belief feels different from a desire because it is generated by different sub-personal processes, which are themselves generated by differential activations induced by our perception of the other.

[8] There is now good scientific evidence to suggest that infants are indeed capable of understanding the intentions of others. Meltzoff (1995) shows that children at 18 months of age are capable of recognizing and completing another person's failed intention. The experimenter pretends to have great difficulty accomplishing a certain task with a toy, and presents an incomplete action to the child. The child, who takes the toy and completes the task with little effort, demonstrates that she understands what the experimenter desired to do. Even earlier, infants seem capable of parsing intentions (see Baird and Baldwin 2001; Baldwin and Baird 2001). Whether some animals are capable of understanding intentions is still a debated issue.

phenomenological argument. On the explicit version of ST, simulation is not only explicit but pervasive. That is, we use it all the time, or at least it is the default way of understanding others. Goldman (2002: 7-8) thinks this is a moderate claim.

> The strongest form of ST would say that all cases of (third-person) mentalization employ simulation. A moderate version would say, for example, that simulation is the *default* method of mentalization … I am attracted to the moderate version …. Simulation is the primitive, root form of interpersonal mentalization.[9]

If simulation is both explicit and pervasive, then one should have some awareness of the different steps that one goes through as one consciously simulates the other's mental states. But there is no phenomenological evidence for this. When I interact with or come to understand another person, there is no experiential evidence that I use such conscious (imaginative, introspective) simulation routines. That is, when we consult our own common experience of how we understand others, we don't find such processes. Of course, this is not to say that we never use simulations; but that in itself is telling. It may be the case that confronted with some strange or unaccountable behavior I do try to understand the other person by running a simulation routine. I think this is the rare case, however. Moreover, it tends to stand out in its rarity. I can easily become aware that I am in fact taking this approach, and it is all the more apparent when I do this, simply because it tends to be the exception. But this tells against the idea that I employ simulation as a default process in the usual everyday circumstance. Most of our encounters are not third-person puzzles solved by first-person procedures. They are second-person interactions in which I easily have a sense of what is going on with the other person based on our common pragmatic or socially contextualized interactions, with no cognitive simulation required.

A possible defensive strategy for theorists of explicit ST is to make it a little implicit. Perhaps explicit simulation can be made so habitual that it becomes implicit, so that we do it without being aware that we do it, in the same way that we drive a car without being explicitly aware of all of our driving habits, or in the same way that an expert may employ cognitive strategies that become so habitual that the expert is no longer aware of how she does what she does. The simple phenomenological objection would be that if such implicit processes stay at the personal level, they would remain accessible to conscious reflection, or at least they would become apparent, as unworkable habits, in problematic situations when our habitual strategies break down. We can become aware of a habit that we are not usually aware of in such circumstances. This simply does not seem to be the case for the sort of simulation process described by explicit ST. Indeed, we may find ourselves initiating simulation processes, sometimes successfully, precisely in cases where our ordinary (non-simulative) abilities to understand others break down.

One could look at the more seriously implicit, i.e., sub-personal or neuronal, version of ST as a way to defeat the simple phenomenological argument. If simulation is sub-personal, and not something of which we would be aware, then phenomenology is

[9] Third-person mentalization signifies simply that one person is trying to understand another person, rather than trying to understand herself (which would be first-person). Some theory-of-minders contend that we use simulation (or theory) to understand our own minds.

not in a position to raise objections, since phenomenology gives us access only to conscious experience. But it is also the case that the appeal to implicit ST actually strengthens the argument against the explicit version of ST. That is, if our understanding of others is in fact mediated by an implicit and automatic simulation process, then we have little need for the more explicit version. Indeed, to the extent that an implicit ST would explain the phenomenological scarcity of explicit simulation, it would support the simple phenomenological argument against explicit simulation. Along this line Gallese states: "Whenever we face situations in which exposure to others' behavior require a response by us, be it active or simply attentive, we seldom engage ourselves in an explicit, deliberate interpretive act. Our understanding of a situation most of the time is immediate, automatic, and almost reflex like" (2005:102).

The sub-personal implicit version of ST is precisely the one that has received the boost from neuroscientific evidence involving sub-personal activation of mirror neurons, shared representations, or more generally, resonance systems. Let's look closely at the supporting neuroscience. The basic finding in this regard is that one's motor system reverberates or resonates in one's encounters with others. My motor system, with its mirror neurons and shared representations, is activated when I perceive another person performing an intentional action, for example (Rizzolatti et al. 1996, 2000; Grezes and Decety 2001). Gallese captures it clearly:

> when we observe goal-related behaviours … specific sectors of our pre-motor cortex become active. These cortical sectors are those same sectors that are active when we actually perform the same actions. In other words, when we observe actions performed by other individuals our motor system 'resonates' along with that of the observed agent …. action understanding heavily relies on a neural mechanism that matches, in the same neuronal substrate, the observed behaviour with the one [the observer could execute] … (2001: 38-39)

These subpersonal mechanisms are said to constitute a simulation of the other's intentions (Gallese 2001; Gallese and Goldstein 1998). Let me focus on Gallese's statements in this regard.

> According to this hypothesis, 'understanding' is achieved by modeling [simulating] a *behaviour* as an *action* with the help of a motor equivalence between what the others do and what the observer does. (39)

This modeling is a subpersonal process effected by "automatic, implicit, and nonreflexive simulation mechanisms …" (Gallese 2005: 117). He refers to this as the "shared manifold hypothesis" and distinguishes between three levels (Gallese 2001: 45):

- The *phenomenological level* is the one responsible for the sense of similarity … that we experience anytime we confront ourselves with other human beings. It could be defined also as the *empathic* level ….
- The *functional level* can be characterized in terms of simulation routines, *as if* processes enabling models of others to be created.

- The *subpersonal level* is instantiated as the result of the activity of a series of mirror matching neural circuits.

Implicit ST nicely avoids the kind of objection raised by Scheler and Ryle, that is, that since simulation is first-person, or at lease confined to my own system (a simulation in my own mind or motor system) nothing justifies inferring anything about what must be going on in the other person. Implicit ST answers this by taking mirror neuron activity (and shared representations), and therefore the core of the simulation routine, as *neutral* in regard to agent identification – such activation represents neither first nor third person agency, as we saw above. On this reading, however, the sub-personal simulation process, like its explicit cousin, involves a multi-step process. On one version of this process, we first perceive the other's behavior; this is followed immediately by activation of mirror neurons or shared representations -- in neutral mode (signifying "naked intentions"); and this is followed by a determination of agency (i.e., a specification of who did the action – me or the other person), and all of this is put in the service of understanding the other person's actions.

Gallese and Goldman (1998) develop a similar model. Mirror neuron activation represents a plan of action. When mirror neurons are activated by observation of the other's movement, the attributor consults this motor plan. "[T]his seems to be a 'primitive' use of simulation with the same structure as [X]", where X is equivalent to a retrodictive use of simulation. "After observing the target agent **(T)** perform action **m**, the attributor uses simulation to test whether goal **g** would have fitted with the choice of **m**. Goal **g** is re-created and fed into his decision-making system, which does output **m**" (p. 498). Where the line is between the personal and the sub-personal levels is not clear, but it does seem clear that Gallese and Goldman want to posit a structural isomorphism between these simulation systems.[10]

In these hybrid models of simulation we come back to the fallacy of supposed isomorphism. At the very least, for both the Gallese-Goldman model and the Jeannerod-Pacherie model, what makes the activation of mirror neurons a simulation process is that in the same way that in mental (explicit) simulation I use my own mental processes to model the other person's mental processes, the neural processes that are used to plan my own actions are employed to model the other person's actions. But these hybrid theories go further in positing simulation as a step-wise process on both levels. Simulation, according to various versions of ST, involves a step-wise process that begins with perception and ends with some form of understanding. We first see an action that we need to understand; we then simulate it in our own mind or motor system; we then infer or understand something about the other's experience. As we have seen, in their hybrid theory, Jeannerod and Pacherie make attribution of agency a separate component, and they maintain that the perceiving of naked intentions and then the deciding about to whom they belong is an articulation that happens on both the neurological and phenomenological levels.

Given other things that they say in their paper, I find it hard to believe that Jeannerod and Pacherie truly mean this, i.e., that there is a phenomenal articulation

[10] If Gallese and Goldman stop short in claiming that mirror neurons constitute a "full-scale realization of the simulation heuristic," the less hybrid models of implicit simulation do seem to make this claim.

between perceiving an action and attributing agency, it is what they say.[11] Of course we could invoke the principle of charity here, and simply ignore what they say. No one likes to be ignored, however, and taking their statements at face value is much more interesting. To be fair, let me point to something in their paper that I can agree with. Indeed, I think that Jeannerod and Pacherie (2004) nicely express the phenomenological alternative to ST: "Perception and action are closely integrated and when we visually perceive actions, we seem to be immediately sensitive to the distinctive properties of intentional behavior" (p. 139). Indeed, they go on to argue for this kind of position.

This is just the way that I would want to put it, without denying the real articulation of processes at the neuronal level. My claim is that the neural processes of which we have been speaking, are part of the processes that underlie intersubjective *perception* rather than *simulation*. That is, these processes have to do with a direct perception of the other person's intentions, rather than a distinct mental process of simulating their intentions.[12] This claim requires that we conceive of perception as a temporal phenomenon, and as enactive, and thus involving motor processes.

First, mirror neurons fire 30-100 ms after appropriate visual stimulation (Gallese, private correspondence). This short amount of time (short certainly in phenomenological if not in neurological terms) between activation of the visual cortex and activation of the pre-motor cortex, raises the question of where precisely to draw the line between the act of perception and something that would count as a simulation. Even if it is possible to draw a line between activation of the visual cortex and activation of the pre-motor cortex, this does not mean that this line distinguishes, on either a functional or phenomenological level between perception and simulation as a step-wise process. That is, even if neuronal processes that send information from sensory cortex to pre-motor cortex take some time (as much as 100 ms), it is not clear that we should identify two functionally distinct steps rather than a temporally extended and enactive perceptual process. At least in terms of temporal parameters, the fact that at the neurological level sensory processing is followed by activation of mirror neurons does not mean that at a functional level one should distinguish between perception and simulation, or that at a phenomenological level there is a similar step-wise process happening.

Furthermore, if we think of perception as an enactive process (e.g., Noë 2004; Noë and O'Regan 2002; O'Regan and Noë 2001; Hurley 1998) – as sensory-motor, and not just as sensory input – then it may be more appropriate to think of resonance processes as part of the structure of the perceptual process when perception is of the action of conspecifics or members of close species. Mirror activation is not the initiation of simulation; it is part of a direct perception of what the other is doing. At the phenomenological level, when I see the other's action or gesture, I see (*directly 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. I see it. I don't have to simulate it. And I immediately see that it is *their* action, gesture, emotion, or intention,

---

[11] One can find a similar articulation in top-down models of agency attribution, as in Graham and Stephens' (2000) explanation of schizophrenic thought insertion and delusions of control. For them, the sense of agency is not a first-order phenomenal aspect of the experience of action, but the result of a higher-order attribution process, which is introspective in the case of thought. For critical discussion, see Gallagher (in press –c).

[12] This view, by the way, is consistent with phenomenologists like Scheler, and Merleau-Ponty (who cited Scheler on this, see, e.g., 1966, p. 352).

and it is extremely rare that I would be in a position to confuse it with my own.

Of course the simulationist can accept the phenomenology ("Yes indeed, that is what *seems* to happen") and still hold to the interpretation that these specific subpersonal processes involve simulation. In fact, however, if we use Gallese's distinctions, this interpretation goes from the functionalist level to the neurological level, but with no clear justification. That is, what happens on the neurological level is simply a complex sequence of neuronal activations. If we look at those processes from a functionalist perspective already framed by ST, then we tend to read those processes as involving simulation. If, in contrast, we look at those processes from a phenomenological level that suggests a direct perception of the other's intentions, then we tend to read those processes as perceptual without simulation. Can the simulationist offer any convincing evidence that the activation of resonance processes is in fact a simulation?

This brings us to ask about the very concept of simulation. I propose that what theorists of implicit simulation (Gallese, Jeannerod, Pacherie), and even critics of implicit ST, like Rebecca Saxe (2005), call 'simulation', is not simulation in any genuine sense of the word. Consider, first, two definitions of 'simulation' offered by the OED.

(1) Simulation is an imitation, in the sense of something *not real* – counterfeit; to simulate means to feign, to pretend. We can find this use of the term in Wittgenstein. "Why can't a dog simulate pain? Is he too honest? Could we teach a dog to simulate pain? Perhaps it is possible to teach him to howl on particular occasions as if he were in pain, even when he is not" (Wittgenstein 1958: *§* 250).
(2) Simulation in the sense of a simulator: a model (a thing) that we can *use* or do things with so we can understand the real thing.

We can find both senses of the term in the literature of ST. Consider the following characterizations (italics are mine). Simulation means "*using* one's own evaluation and reasoning mechanisms as a model for theirs …" (Dokic and Proust 2002: viii). Simulation involves "pretend states" where,

> by pretend state I mean some sort of surrogate state, which is *deliberately adopted* for the sake of the attributor's task … In simulating practical reasoning, the attributor *feeds* pretend desires and beliefs into her own practical reasoning system. (Goldman 2002: 7).

The surrogation or pretense however, is of a precise kind. Bernier (2002) makes this explicit as an essential element found in ST.

> According to ST, a simulator who runs a simulation of a target would *use* the resources of her own decision making mechanism, in an "off-line" mode, and then the mechanism would be fed with the mental states she would have *if she was in the target's situation*. (Bernier 2002: 34)

For ST, a simulation is not simply a model that we use to understand the other person -- theoretical models would suffice if this were all that is required. Even the fact that the

model is constituted in our own mechanisms is not sufficient. Rather, I must use the model "as if" I were in the other person's situation. As Gallese puts it, "our motor system becomes active *as if* we were executing that very same action that we are observing" (2001: 37). Gordon locates this "as if" right at the neuronal level: the neurons that respond when I see your intentional action, respond "*as if* I were carrying out the behavior …" (2005: 96). If we call this the "pretense condition," then simulation has these two characteristics: it is a process that I control (hence all of the action words in the above characterizations, and in the explicit version it is "deliberately adopted"), and it involves a pretense condition (I put myself "as if" in the other person's shoes).

We find these two characteristics in almost every description of simulation in ST. Let's look again. Mental simulation is a cognitive "*ability* or *heuristic* or *methodology"* (Jacob 2002, who cites Gordon for the latter term) -- by which I "*engage* in pretense," *put myself* in someone else's shoes, *compare* my experience to their experience, and *predict* their mental state, emotion, or behavior. We *use* ourselves as a model … I *create* in myself some pretend beliefs … and so forth. This is the way simulation is characterized not only by theorists of explicit simulation, but also by theorists of implicit simulation. The pretense condition mentioned by Gallese is accomplished in a simulation considered to be "an interactive *model* of what cannot be known in itself" (2003). At the subpersonal level, the brain in a stepwise fashion is *modelling* the intentional action of others. Gordon (2004: 1) suggests that on the "cognitive-scientific" model, "one's own behavior control system is employed as a manipulable model of other such systems. (This is not to say that the "person" who is simulating is the model; rather, only that one's brain can be manipulated to model other persons)."

If simulation is characterized as a process that I (or my brain) uses or controls, if this is what simulation is, then it seems clear that what is happening in the implicit processes of motor resonance is not simulation. We, at the personal level, do not *do* anything with the activated brain areas -- in fact, we have no access to neuronal activation, and we can't use it as a model. Nor does it make sense to say that at the subpersonal level the brain is *using* a model or methodology, or *comparing* one experience with another, or *creating* pretend states, or that one set of neurons makes use of another set of neurons as a model.[13] I think that would put us back very close to Sandra Blakeslee's phrasing: that mirror neurons read minds and understand actions. Rather, we must say that neuronal systems do *not* take the initiative; they do *not* activate themselves; they are activated by the other person's action. The perception of the other person's action automatically activates in our brain the same areas that are activated when we engage in similar action. The other person *has an effect on us*. The other *elicits* this activation. This is not a simulation, but a perceptual elicitation. It is not us (or our brain) *doing* it, but the other who does this to us.[14]

---

[13] I'm not sure the problem is solved by attributing such things to the simulation itself, as Hurley (2005) does: "Simulation *uses* certain processes to generate related information, rather than *theorizing about* them in separate meta-processes."

[14] It may seem contradictory to claim in the previous argument that perception is enactive, or as Noë says, "perception is action," and in this argument to claim that the activation of the resonance system is the result of a passive elicitation, so that the motor aspect of perception does not involve our action, but is a case of us being affected by the other. I think that a fuller account of enactive perception has to be able to accommodate this passive, affective aspect of perception (see Gallagher 2005). It would be interesting to link this idea of elicitation to a certain ethical dimension that goes beyond the concept of empathy, along

But wait. Even if in the case of intersubjective perception resonance processes described as the activation of mirror neurons and shared representations are not initiated and controlled by my brain or by me, there is another way to think of simulation on the subpersonal level that does involve controlled comparison. In the motor control literature and in discussions of agency, theorists do talk about comparators and models used by the brain. The term 'simulation' is also used. Efference copy sent through forward control mechanisms, for example, are said to constitute a simulation of an intended movement in order to compare it with an ongoing movement, to predict its success (e.g., Berthoz 2000; Wolpert et al. 1995). The brain uses this simulation to make fast non-conscious corrections that keep the action on track. This use of the term seems closer to its general meaning even though these processes are automatic. In regard to motor control, the use of the term 'simulation' may be acceptable, and is certainly better than the term 'representation' (see Berthoz 2000, 22). In addition, the forward-control model helps to make the close connection between perception and action. Some theorists, however, have appealed to these processes as possible mechanisms involved in the simulation of another's action, i.e., specifically as a mechanism that serves theory of mind (e.g., Gallese 2001; Hurley 2005: 181-188; see Iacoboni, cited in Millikan 2005: 188n2). Our own motor system comparators are activated to simulate and thereby anticipate the other's action. The brain could be said to predict the other person's actions in this way. On this account the perception of the other's action is automatically informed by a sub-personal simulation; perception of action involves a loop through the motor control comparator. Can ST adopt this model of simulation – does this use of the term conform to the ST concept?

The problem in this case is that the pretense condition is not met; there is no "*as if* it were I" involved, and in that regard it fails to be the kind of simulation required by ST. If indeed the subpersonal simulation is neutral in regard to who's action is at stake (Gallese 2005; Jeannerod and Pacherie 2004; Hurley 2005), then it can be only a representation of an intentional action in my motor system, but not a representation of my own motor action as if it were the other's. Even if a "Who" mechanism adds a specification of agency, differentiating between self and non-self, it is not at all clear that, as Gordon (2005: 96) suggests, the neurons respond "*as if* I were carrying out the [other's] behavior" in any sense in which the "as if" registers sub-personally.[15] A

---

the lines of Levinas, or possibly Jean-Luc Marion's (2002) notion of a calling or seduction. To pursue these ideas would be to go far beyond the lines of this paper, but these ideas do suggest how limited the concept of simulation is.

[15] It is possible to use brain imaging to identify neuronal areas differentially activated for first-person simulated action (the subject imagines herself doing the action) and third-person simulated action (the subject imagines that she is the other person doing the action) (Ruby and Decety 2001). Results show that there is a good deal of overlap of neuronal areas responsible for these different activities, but also some non-overlapping areas that supposedly accounts for the differences in agency and perspective. Similar studies show overlapping areas for first-person action and first-person simulated (imagined) action. Thus, although it seems reasonable that one might be able to point to a certain area of neuronal activity associated with pretense (imaginative enactment or simulation) in these cases, such activation is generated by a personal level, consciously controlled pretense. What has not been shown, as far as I know, is an automatic pretense condition that takes place without conscious control, on the subpersonal level. Indeed, the concept of pretense, the "as if," may simply be a personal level category that cannot be applied to neuronal activity *per se*. Although one might say that neurons respond "as if" X were the case (e.g., *as if* there were an oasis,

specification in my motor system that the action belongs to another is not equivalent to the specification "*as if* I were carrying out the action." If this is a simulation of intentional action, it is nonetheless not the kind of simulation that ST needs; it may be nothing more than motor priming or emulation, or what Hurley calls mirroring (2005: 184).

It is therefore not clear why we should think of the activation of resonance systems as a simulation process of the sort required by ST. This is not to deny that there are resonance processes at work in our perception of the other person. Moreover, the nature of the resonance processes involved in such encounters makes our perception of other conspecifics different from our perception of objects and instruments. But it doesn't make social cognition the result of an implicit simulation.

Conclusion

I have brought a number of arguments against various versions of ST. What I've called the simple phenomenological argument can challenge explicit ST, especially if we take these versions to describe our primary, pervasive, or default mode of understanding others. In such cases, ST should be able to call on phenomenological evidence to verify the simulation model. In fact, phenomenological reflection on our ordinary experience speaks against ST. In our everyday encounters we find a scarcity simulation routines at work. One could argue, of course, that phenomenology is not always correct, even in regard to the most explicit experiences. But that argument would have to be extended too far if the claim were to be made that in fact we use simulation routines all the time, but phenomenology constantly and consistently misses this fact.

In addition to what I've called the simple phenomenological argument, phenomenologists and others, like Ryle, have pointed to logical and conceptual problems with ST. Given the large diversity of motives, beliefs, desires, and behaviors in the world, it is not clear how a simulation process based on my own relatively narrow experience (or relatively unique circumstance) can give me a reliable sense of what's going on in the other person's mind, or in their behavior.

Implicit ST may seem to be in a better position to answer this particular charge. The problem with implicit ST is that it is only one, and not necessarily the best interpretation of the significance of motor resonance systems. We noted that if implicit ST does give an account of our primary and pervasive ability to understand others, it would count as an argument against explicit ST, since explicit simulations would be redundant in this case. Likewise, however, if our default mode of understanding others were based on explicit simulation, as Goldman contends, then the claims of implicit ST about the adequacy of motor resonance processes would be wrong. Goldman's view of implicit motor resonance processes is that they do not constitute simulations of a sort that would be sufficient to do the full job, but do generate some background information that is useful to initiate the explicit simulation process. If a hybrid ST is successful in combining these otherwise antagonistic versions, and if it can escape the charge that either the explicit or the implicit side is redundant, it seems to run into the fallacy of supposed isomorphism.

---

even though it is a mirage) it is not clear that the "as if" aspect is neuronal rather than an interpretation of that activity.

I've argued that implicit resonance processes are not simulations in any sense that is useful for ST. In contrast to the claims made by ST, implicit motor resonance processes are important enactive processes that contribute to the constitution of the perceptual access that we have to the intentions of others. If it makes any sense to talk about naked intentions at the subpersonal level, at the phenomenological level all intentions are fully clothed with agency. I do not claim that we get a full account of human intersubjectivity in the idea that we have *perceptual* access to the intentions of others. Perceptual access to the other person's bodily movements, gestures, facial expressions, and so forth does give us a sense of what is going on with them, what they mean and what they feel, and this, together with our interactions with them in pragmatic and social contexts, gives us a relatively reliable, but still minimal understanding of them. There is much more to say about the role of language and narrative competency in a fuller account of intersubjectivity (see Hutto 2003, 2004, and forthcoming; Gallagher in press, Gallagher & Hutto forthcoming). Even in that larger story, however, theory of mind approaches that emphasize simulation (or the role of folk psychology as background theory), have a minimal role to play in our normal and everyday interactions.

References

Baird, J. A. and Baldwin, D. A. 2001. Making sense of human behavior: Action parsing and intentional inference. In *Intentions and Intentionality: Foundations of Social Cognition*, ed. B. F. Malle, L. J. Moses, and D. A. Baldwin (pp. 193-206). Cambridge, MA: MIT Press.

Baldwin, D. A. and Baird, J. A. 2001. Discerning intentions in dynamic human action. *Trends in Cognitive Science* 5 (4): 171-78.

Bernier, P. 2002. From simulation to theory. In J. Dokic and J. Proust (eds.), *Simulation and Knowledge of Action* (33-48). Amsterdam: John Benjamins.

Berthoz, A. 2000. *The Brain's Sense of Movement*. Cambridge, MA: Harvard University Press.

Currie, G. and Ravenscroft, I. 2002. *Recreative Minds*. Oxford: Oxford University Press.

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

Dokic, J. and Proust, J. 2002. Introduction. In J. Dokic and J. Proust (eds.), *Simulation and Knowledge of Action* (vii-xxi). Amsterdam: John Benjamins.

Evans, G. 1982. *The Varieties of Reference*, ed. J. McDowell. Oxford: Oxford University Press.

Fadiga, L., et al. (1995), 'Motor facilitation during action observation: a magnetic stimulation study', *Journal of Neurophysiology*, 73, 2608-2611.

Farrer, C. and Frith, C.D. 2002. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage* 15 (3), 596–603.

Farrer, C., Franck, N., Georgieff, N., Frith, C.D. Decety, J. and Jeannerod M. 2003. Modulating the experience of agency: A positron emission tomography study. *NeuroImage* 18: 324–333.

Gallagher 1997. Mutual enlightenment: Recent phenomenology in cognitive science. *Journal of Consciousness Studies* 4 (3): 195-214

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. (in press -a). Logical and phenomenological arguments against simulation theory. In D. Hutto and M. Ratcliffe (eds.), *Minding our Practice: Folk Psychology Re-assessed*. Springer Publishers.

Gallagher, S. (in press -b). The narrative alternative to theory of mind. *Consciousness and Emotion.*

Gallagher, S. 2005. Book review: A new movement in perception: Review of Alva Noë's *Action in Perception*. *Times Literary Supplement* (London).

Gallagher, S. and Hutto, D. (forthcoming). Primary interaction and narrative practice: From empathic resonance to folk psychology. In: Zlatev, Racine, Sinha and Itkonen (eds). *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins.

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

Gallese, V. 2003. The Roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36: 171-180.

Gallese, V. 2001. The 'shared manifold' hypothesis: from mirror neurons to empathy', *Journal of Consciousness Studies* 8: 33-50.

Gallese, V., & Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2: 493-501.

Georgieff, N. and Jeannerod, M. 1998. Beyond consciousness of external events: A 'Who' system for consciousness of action and self-consciousness. *Consciousness and Cognition* 7: 465-77.

Goldman, A. I. 2002. Simulation theory and mental concepts. In J. Dokic and J. Proust (eds.), *Simulation and Knowledge of Action* (1-19). Amsterdam: John Benjamins.

Goldman, A. 2005a. Mirror Systems, Social Understanding and Social Cognition. *Interdisciplines.* (http://www.interdisciplines.org/mirror/papers/3).

Goldman, A. 2005b. Imitation, mind reading, and simulation. In Hurley and Chater (eds.) *Perspectives on Imitation* II (80-81). Cambridge, MA: MIT Press.

Gordon, R. M. 2004. Folk psychology as mental simulation. In N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. (URL = http://plato.stanford.edu/archives/fall2004/entries/folkpsych-simulation/).

Gordon, R. M. 1995a. Folk Psychology as Simulation. In: M. Davis & T. Stone (Eds.). Folk Psychology. Oxford: Blackwell. 60–73.

Gordon, R. M. 1995b. Developing commonsense psychology: Experimental data and philosophical data. APA Eastern Division Symposium on Children's Theory of Mind, December 27, 1995. (http://www.umsl.edu/~philo/Mind_Seminar/New%20Pages/papers/Gordon/apakids9.htm).

Gordon, R. 1986, Folk psychology as simulation. *Mind and Language* 1: 158-171.
Gordon, R. and Cruz, J. 2003. Simulation theory. *Encyclopedia of Cognitive Science* London: The Nature Publishing Group Macmillan Reference Ltd.
Grèzes, J. and Decety, J. (2001), Functional anatomy of execution, mental simulation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1-19.
Heal, J. 1986. Replication and functionalism. In J. Butterfield (ed.), *Language, Mind, and Logic*. Cambridge: Cambridge University Press.

Hurley, S. L. 2005. Active perception and perceiving action: The shared circuits model. In T. Gendler and J. Hawthorne (eds.), *Perceptual Experience*. New York: Oxford University Press.
Hurley, S. L. 1998. *Consciousness in Action.* Cambridge, MA: Harvard University Press.
Hutto, D. 2003. Folk-psychological narratives and the case of autism. *Philosophical Papers* 33 (3): 345-61.
Hutto, D. 2004. The limits of spectatorial folk-psychology. *Mind and Language* 19: 548-73.
Hutto, D. (in press). Unprincipled engagements: Emotional experience, expression and response.
Hutto, D. (forthcoming). *Folk Psychological Narratives*. Cambridge, MA: MIT Press.
Jacob, P. 2002. The scope and limits of mental simulation. In J. Dokic and J. Proust (eds.), *Simulation and Knowledge of Action* (87-109). Amsterdam: John Benjamins.
Jeannerod and Pacherie 2004. Agency, simulation, and self-identification. *Mind and Language* 19 (2) 113-46.
Marion, J-L. 2002. *Being Given: Toward a Phenomenology of Givenness.* Trans. J. Kosky. Stanford: Stanford University Press.
Meltzoff, A.N. 1995. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology* 31: 838-850.
Millikan, R. G. 2005. Some reflections on the theory theory-simulation theory debate. In Hurley and Chater (eds.) *Perspectives on Imitation* II (182-88). Cambridge, MA: MIT Press.
Mitchell, J. P. 2005. The false dichotomy between simulation and theory-theory: the argument's error. *Trends in Cognitive Sciences* 9 (8): 363.
Nichols, S. and Stich, S. 2003. *Mindreading*. Oxford: Oxford University Press.
Noë, A. 2004. *Action in Perception*. Cambridge, MA: MIT Press.
O'Regan, J. K. and Noë, A. 2001. A sensorimotor approach to vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5): 939-73.
Rizzolatti, G., Fogassi, L. and 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. 539-52). Cambridge, MA: MIT Press.
Rizzolatti, G., Fadiga, L., Gallese V. and Fogassi, L. 1996. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3: 131-141.
Ruby, P. and Decety, J. 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, 4 (5): 546-50.
Ryle, G. 1949. *The Concept of Mind.* New York: Barnes and Noble.

Scheler, M. 1954. *The Nature of Sympathy*. Trans. P. Heath. London: Routledge and Kegan Paul. Original: *Wesen und Formen der Sympathie*. Bonn: Verlag Friedrich Cohen, 1923.

Saxe, R. 2005. Hybrid vigour: Reply to Mitchell. *Trends in Cognitive Sciences* 9 (8): 364.

Varela, F. J. 1996. Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies,* 3 (4): 330-49.

Wittgenstein. L. 1958. *Philosophical Investigations*. Trans. G. E. M. Anscombe. Third edition. New York: Macmillan.

Wolpert D.M., Ghahramani Z. and Jordan M.I. 1995. An internal model for sensorimotor integration. *Science* 269:1880-1882

Zahavi, D. 2004. Back to Brentano? *Journal of Consciousness Studies* 11 (10-11): 66-87.

Zahavi, D. 1998. Brentano and Husserl on self-awareness. *Études Phénoménologiques* 27-28: 127-168