Gallagher, S. 2007. Neurophilosophy and neurophenomenology. In L. Embree and T. Nenon (eds.), *Phenomenology 2005* Vol. 5. (293-316). Bucharest: Zeta Press.

Neurophilosophy and neurophenomenology[1]

Shaun GALLAGHER
Philosophy and Cognitive Sciences
University of Central Florida
gallaghr@mail.ucf.edu

ABSTRACT: The neurophilosophical project, as envisioned by Churchland, involves intertheoretic reduction, moving from (or eliminating) theories formulated in terms of common sense and folk psychology, to theories that have stood the test of scientific experiment. In her view, folk psychology, as well as introspective phenomenology, will be eliminated in favor of neuroscience. Neurophenomenology holds that phenomenology (as a practice) is not only possible, but is in fact a useful tool for science; and that phenomenology is ineliminable if the project is to pursue a neurobiology of consciousness. Clarification of these issues rests on an understanding of how phenomenology can be an alternative source of testable theory, and can play a direct role in scientific experiment. Rather than talking in the abstract about the role of theory formation in science, I consider two specific issues to show the difference between a neurophilosophical approach and a neurophenomenlogical approach, namely, the issues of self and intersubjectivity. Neurophilosophy (which starts with theory that is continuous with common sense) and neurophenomenology (which generates theory in methodically controlled practices) lead to very different philosophical views on these issues.

---

[1] This is the text of a plenary paper presented at the conference: *Toward a Science of Consciousness 2005 – Methodological and Conceptual Issues* (August 17-20, 2005), Copenhagen, Denmark.

> One could speak of Socrates, to wonder whether he was a good philosopher or a bad one; but one can also speak of sleep or of boredom, and if one cannot say much about [the word] 'and' … there remains much to say about the whole ensemble of processes, of manners of being, of actions, of sensations, or of impressions that one cannot consider as objects. … [I]t seems to me … that one can legitimately pose some question about – let us say – memory; in what does memory consist? Is it essential to reserve this notion to designate only those experiences that are our own? …. And it is not impossible that this is the genre of research that certain disciples of Husserl recommend, in which case their curiosity seems to me perfectly legitimate (A. J. Ayer 1960).

Getting all Ryled up

In this passage A. J. Ayer was in the process of posing a question to Gilbert Ryle about whether it is possible to be concerned about experience, perhaps in the way that Husserl's phenomenology is concerned about experience, even if we take ordinary language statements as our starting point. One might be surprised to find two analytic philosophers, who were actually discussing an argument made by Wittgenstein, suddenly evoking Husserl's phenomenology. But, as Wittgenstein himself might say, context is everything. The setting for this discussion was a meeting at Royaumont which put Ryle, Ayer, and Quine into conversation with R. P. Van Breda (the original founder of the Husserl Archives) and Merleau-Ponty. Merleau-Ponty picked up where Ayer left off:

> I have also had the impression, while listening to Mr. Ryle, that what he was saying was not so strange to us [phenomenologists], and that the distance, if there is a distance, is one that he puts between us rather than one I find there (Merleau-Ponty 1992).

Mr. Ryle goes on to agree with Ayer that philosophers can legitimately talk about memory, perception, sleep, and other ambiguous non-objects, but he does not recant a criticism he had made of Husserl, namely that to the extent that Husserl had talked about meaning that was not reducible to verbal meaning, his descriptions were nonsense. Perhaps more surprising than his dismissal of phenomenology from the realm of philosophy, which he seemingly limits to conceptual analysis, is his dismissal of empirical fact. After describing examples of what empirical scientists do, he states: "See here what comes to my mind when speaking of research of fact. Nothing very mysterious, as you see. But what matters is that the questions of fact of this order are not the province of philosophy. One will never say that so and so is a better philosopher than so and so because so and so knows facts of which the other is ignorant" (Ryle 1960). [2]

Times have changed in some ways; and in some ways not. Would Dan Dennett or Patricia Churchland agree with Ryle about a philosopher's knowledge of empirical research? Churchland follows Quine rather than Ryle: "philosophy at its best and properly conceived is continuous with the empirical sciences" (1986: 2). But just as much as they disagree with Ryle on this point, they agree about his view of phenomenology (despite Merleau-Ponty's wishful thinking). Dennett writes, in response to his own encounters with phenomenologists, and with a curiously 19th-century use of capitals:

> I studied Husserl and the other Phenomenologists with Dag Føllesdal at Harvard as an undergraduate, and learned a lot. My career-long concentration on intentionality had its beginnings as much with Husserl as with Quine. But part of what I thought I learned from those early encounters is that reading the self-styled Husserlians was largely a waste of time; they were deeply into obscurantism for its own sake. I may have picked this attitude up from my graduate advisor, Gilbert

[2] Quotes from Ayer, Ryle, and Merleau-Ponty are from Merleau-Ponty 1992: 63-65.

> Ryle, who was himself a masterful scholar of Husserl and Phenomenology. In any case, when we discussed my own work on intentionality he certainly didn't encourage me to follow him in attempting to plumb the depths of the Continental Husserlians (Dennett 1996).[3]

Dennett goes on to tell of the poor reception he received in Paris and Nice. "The French Husserlians either were aghast or found me beneath notice, in spite of my attempt to convey my sense of my Husserlian heritage." I fear that much of this is tongue-in-cheek, but trust that his great praise for Eduard Marbach is dead serious ("I take very seriously Eduard Marbach's recent and forthcoming attempts to build a bridge

[3] Dennett's view of Ryle as well versed in phenomenology is not off the mark. As Amie Thomasson (2002) notes, "historical evidence of Ryle's serious and substantial interest in this tradition is abundant; indeed his masterfully concise fifteen-page autobiography devotes a generous two full pages to recounting his studies of Husserl and the phenomenologists. (By comparison, the Vienna Circle and Wittgenstein each merit less than a page). His first publication ever is a review of (Husserl's student) Roman Ingarden's *Essentiale Fragen* (1927); his second is a review of Heidegger's *Being and Time* (1929). Early in his career as a don at Oxford, as Ryle reports, he "offered an unwanted course of lectures, entitled 'Logical Objectivism: Bolzano, Brentano, Husserl and Meinong'", … In 1929 he traveled to Freiburg where he "spent an hour discussing phenomenology" with the aging Husserl, and stayed on to study with Heidegger. In 1932 he (with others) held a symposium of the Aristotelian society on phenomenology. Finally, lest Ryle's interest in phenomenology be (as usual) written off as a youthful indiscretion, it is worth noting that over the course of his lifetime he published at least six essays (spanning most of his career—from 1927 to 1962) focused entirely on the phenomenological tradition and providing it a largely sympathetic exposition to introduce it to the English-speaking world. In the volume of collected papers dedicated to his critical essays on other thinkers … four of the twenty essays included are on phenomenology, a number exceeded only by essays on the ancients, and far more than on any other twentieth century thinker or movement (by comparison, there are two essays on Wittgenstein, one on Carnap and one on Moore). In his topical essays, too, such figures as Brentano, Meinong and Husserl are frequently discussed …"

between my heterophenomenology and (his refreshingly clear version of) Husserl's autophenomenology.")

Certainly, however, for the Post-Rylean philosophers of mind and the neurophilosophers, such as Dennett and Churchland, empirical research has become central. Is it possible, however, that despite absolutely everything they think about and against phenomenology, the importance that they give to empirical research actually comes along with a necessary turn to phenomenology? I am motivated to ask this question in light of recent proposals that empirical research on consciousness requires some form of phenomenological method. This is at least one claim made by Francisco Varela under the heading of neurophenomenology.

There are obvious and important differences between phenomenological approaches and the neurophilosophical approach in the analytic tradition. One of these differences concerns the proper starting point. Patricia Churchland considers the generation of testable theory to be one of the tasks of neurophilosophy. What are the sources of scientific theory? Churchland, following Quine, suggests "Scientific theories … are *continuous* with common sense, they are common sense subjected to critical analysis" (2002: 111). The neurophilosophical project, as envisioned by Churchland, involves intertheoretic reduction, moving from (or eliminating) theories formulated in terms of common sense and folk psychology, to theories that have stood the test of scientific experiment. In her view, folk psychology, as well as introspective phenomenology, will be eliminated in favor of neuroscience.

Notwithstanding certain methodological confusions of phenomenology with introspection or folk psychology, and sometimes of neurophilosophy with neurophenomenology – e.g., Metzinger [2003: 83] writes, clearly in the spirit of neurophilosophy, "Neurophenomenology is possible; phenomenology is impossible" – I say that despite all of this, neurophenomenology holds that phenomenology (as a practice) is not only possible, but is in fact a useful tool for science; and that phenomenology is ineliminable if the project is to pursue a neurobiology of consciousness. Clarification of

these issues rests on an understanding of how phenomenology can be an alternative source of testable theory, and can play a direct role in scientific experiment (see Gallagher 2003).

Let me briefly indicate how I am not going to proceed. First, I am not going to try to make the clarification by showing how phenomenological method can inform experimental practice and design (I've tried to do that elsewhere – Gallagher 2003; Gallagher and Sørensen 2006). Second, I am not going to draw out the philosophical debate about the proper grounds for theory formation, which in the hands of a better philosopher than I am, could easily fill a multitude of pages in a philosophy of science journal. Rather I will consider two specific issues that I think show the difference between a neurophilosophical approach and a neurophenomenological approach, namely, the issues of consciousness and intersubjectivity. My purpose is not primarily to show that neurophilosophy (which starts with a critique of common sense) and neurophenomenology (which suspends common sense in methodically controlled examinations of experience) lead to very different philosophical views on these issues – although I think they do. Rather, my primary purpose is to keep the following question in the foreground: What are we trying to explain when we try to explain consciousness or intersubjectivity? And my suggestion is that neither neurophilosophy nor the empirical sciences can afford to ignore phenomenology or neurophenomenology.

Consciousness

I'm going to use Patricia Churchland as the main representative of neuro-philosophy – and I don't think I need to justify that. But I'll take neurophilosophy in a wide sense to include the work of many analytic philosophers of mind who make strong appeals to neuroscience. In a chapter on consciousness in her recent book entitled *Brain-Wise: Studies in Neurophilosophy* (2002) Churchland is clearly looking for a neurobiological explanation of consciousness. Although, as

she admits, there is no such thing, as of yet (in part because of the immaturity of neuroscience), in her review of possible and promising candidates and strategies, she alights on a proposal made by Antonio Damasio (a neuroscientist, as you know) but also supported by a number of philosophers, and she names David Armstrong, and our friends Thomas Metzinger and David Rosenthal, as well as Paul Churchland. As Damasio explicates it, consciousness involves self-representation, where self means the sensori-motor coordination control mechanisms that can be explained in terms of forward models, but also involving perception and emotion circuits. In the framework of neuroscience this involves the evolution of "new circuitry" that enables one neuronal population to represent the internal model that we use for motor coordination.

> It could represent some items in the model (themselves representations) as standing in relation to representations of states of the body. That is the circuitry could represent certain of the organism's current perceptual and emotional states *as* states of itself, it could categorize some representations as being *of* objects external to the body, and, most important, it could represent the relation between them. (Churchland 2002: 165).

This metarepresentation of lower order representations "enables second-order evaluation structures and second-order planning and predictive structure" (p. 165). This metarepresentational "upgrade" not only grants a survival advantage, it constitutes consciousness as specifically something that emerges in specialized but widely distributed neuronal populations, and in functions that allow for self-attribution ("This pain is mine"), perspectival self-representation, and so on (p. 166).

This is a higher-order representational theory of consciousness – not precisely a HOT theory, as found in Rosenthal (1993), or a HOP theory as found in Carruthers (1998), but a close first cousin in the neuroscience literature. The meta-representation is not itself conscious, but it makes the first-order processes conscious. The theory "identifies

consciousness of pain, for example, with a representation in the metarepresentational schema" (p. 166). For Churchland, this is nicely reductionistic since the capacity for metarepresentation is just a biological, physical capacity of the brain.

A phenomenological approach to this question is obviously a different kind of animal. We might best capture it by outlining the phenomenological objections to higher-order representational theories of consciousness, and see how they apply to the Damasio-Churchland neuroscientific model. Dan Zahavi (2002; Zahavi and Gallagher 2005; Zahavi and Parnas 1999) and several others have provided a clear, phenomenological critique of the HOT version of this approach, and have outlined the phenomenological alternative. Let's see what the criticism looks like when applied to the Damasio-Churchland theory (DC for short).

According to higher-order representation theories, the higher-order representation takes the first-order representation as an object – as Churchland puts it, it represents the first-order representation. But whatever the notion of representing means, the metarepresentation cannot itself be (transitively) conscious, otherwise this would lead to infinite regress. The regress is easily avoided by accepting the existence of non-conscious mental states. This is precisely the position adopted by the defenders of higher-order theory like Rosenthal (1997: 745), and it is clear that for DC too the metarepresentation remains non-conscious. The phenomenological reply to this solution is rather straightforward, however. The phenomenologists would concede that it is possible to halt the regress by postulating the existence of non-conscious mental or neuronal states, but they would maintain that such an appeal to the non-conscious leaves us with a case of explanatory vacuity. That is, they would find it quite unclear why the relation between two otherwise non-conscious processes should make one of them conscious. Or to put it differently, they would be quite unconvinced by the claim that a state without subjective or phenomenal qualities can be transformed into one with such qualities, i.e., into an experience with first-personal givenness

or mineness, by the mere relational addition of a meta-state having the first-state as its object.

Of course neurophilosophy might insist that this is precisely how things must be, unless we argue for something like consciousness all the way down (as Chalmers might). So let's grant the hypothesis that consciousness emerges in some way from non-conscious neurological processes. It is still not clear what it means for the metarepresentation to take the first-order representation as an object – or to represent it. If this means that the lower-order representations are simply integrated into a more complex higher-order metarepresentation (which would simply be a more complex neuronal activation), and if the claim is that this is simply "a biological fact about the brain," then no explanation of consciousness is given at all. It amounts to saying no more than the brain generates consciousness by activating special metarepresentational neuronal systems. These systems must be special, or "consciousness special," so to speak, because there are innumerable examples of neuronal activations being incorporated into larger more complex neuronal activations without anything like consciousness being at stake.

Furthermore, the precise details of what gets metarepresented are important. The first-order representation is, for example, an activation of neurons that represent the body in relation to the world at a sensory-motor level. The meta-representation then "categorizes" some part of the first-order representation system as representing objects external to the body, or as "standing in the 'belongs to' relation to the self-representation." Churchland introduces plenty of intentional terms into the account – the nervous system is capable of "ranking goals, making behavioral decisions, and evaluating relevant perceptual signals" (p. 164). She wants to talk about qualitative differences between experiences, but admits that exactly how these qualitative differences are generated "should be sorted out as neuroscience proceeds." But in this regard we are motivated to ask our question: What precisely is DC trying to explain? What precisely should neuroscience try to sort out? It sounds like neuroscience has the task of trying to

explain our phenomenal experience – not only how it is generated, but what it is like. But what precise details of phenomenal experience do these representations (neural activations) explain –- what features of consciousness do these representations generate when they generate consciousness?

At the very least this suggests that there is some descriptive phenomenology at stake here, and it motivates us to ask what is the phenomenology, and where does Churchland find it? Are we simply trying to cash out folk psychology in neurological terms?

The phenomenological approach attempts to start with a controlled description of the phenomenon at stake. By controlled description I mean one that is constrained by phenomenological method. To be clear, however, the claim is not that phenomenology can explain how the brain generates consciousness – indeed, phenomenological method (the phenomenological reduction) rules out this kind of task for the phenomenologist -- but it can develop a description of the features of consciousness we are trying to explain. The claim is further that this description will constrain the neurological explanation. In other words, phenomenology is not offering an alternative theory to the neurophilosophical DC model, but would provide some essential descriptions that any theory would have to consider.

For example, DC suggests that consciousness emerges in such a way that while acting I am aware of my body as in relation to the environment. Specifically, a first-order representation that gets metarepresented is, according to Churchland, the forward model or "Grush emulator" that represents the body in relation to the environment. Apparently, a metarepresentation simply makes this piece of sub-personal motor control engineering conscious. But phenomenology suggests that this is not the way my conscious experience is structured, especially with respect to what I am aware of when I am engaged in action. E.g., what precisely am I conscious of when I reach to pick up a glass? Not of my bodily movements in any explicit way. And if in any sense I

am conscious of my arm reaching, what modalities are involved, and what if any sense of agency do I have for this action? Unless I know these details, I don't know what I'm trying to explain, or where or what to look for in the brain.

In other words, what we are trying to explain when we are trying to explain how consciousness is generated is not divorced from the precise nature of the conscious experience. If the phenomenology indicates that we are not conscious of X, or that our consciousness is not X-like, but the neurological explanation inserts X into the picture, so to speak, then the neurological explanation is not explaining consciousness as it is.

The claim here is not that consciousness is simply and necessarily what it seems to be, if what it seems to be is characterized by our natural, everyday folk psychological understanding of it. Rather, phenomenology claims that there is a description of consciousness that gets it right, and we get it right if we approach it in a methodologically controlled way. Neurophenomenological methodology integrates phenomenology and experimental procedures that balance first-person accounts with third-person data, and one can argue that this makes the phenomenological description sharper, more secure, and intersubjectively verifiable. If, as neuroscientists, we are looking in the brain to find precisely what schemas of neural activations are doing what, we need to have some kind of precise phenomenological map of what we are conscious of in any particular situation.

Intersubjectivity

In regard to the issue of intersubjectivity or social cognition, both phenomenology and analytic philosophy of mind are of one mind on the importance of understanding how we understand others. Merleau-Ponty pressed Ryle on this point. He asks Ryle how he would deal with the transformation of first-person propositions into second-person propositions which take the first-person to be another person (Merleau-

Ponty 1992: 68). Ryle responds by remarking on the importance of this problem: "the fact that a subject A speaks to another subject B and that this other can respond to him is one of the most important facts that one can consider when one speaks of 'persons'" (Ryle 1960). We have to look to *The Concept of Mind* (1949), however, to find his views on how this works. Specifically, what he says there can be interpreted either in simplistic behavioristic terms, or in terms that are quite similar to phenomenological insight.

Ryle rejects the Cartesian official doctrine that would lead us to believe that the minds of others are hidden away and inaccessible to us. We do not make "untestable inferences to any ghostly processes occurring in streams of consciousness which we are debarred from visiting," but instead, we attend to "the ways in which those people conduct parts of their predominantly public behaviour. True, we go beyond what we see them do and hear them say, but this going beyond is not a going behind, in the sense of making inferences to occult causes; it is going beyond in the sense of considering, in the first instance, the powers and propensities of which their actions are exercises" (1949, p. 51). Ryle further suggests that if we had to depend on making inferences from a knowledge of psychological laws (a version of what today is called "theory theory"), we would be led to a paradox that if someone actually knew these laws, they could never explain them to anyone else who also didn't already know them. Ryle also rejects what today is called simulation theory, or what in the past was called the argument from (inference by) analogy, on much the same grounds as the early phenomenologists, showing that it would be a fallacious inference: "the 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" (p. 54). More positively, Ryle's answer is reminiscent of Heidegger's view, as expressed in *Being and Time*, a book that Ryle reviewed in *Mind* (1929), as Dennett

would no doubt want to point out.[4] For Ryle, as for Heidegger, understanding others "is part of knowing *how"* – it requires understanding performance in context.

On such points, contemporary neurophilosophers pay no attention to Ryle. Indeed, Churchland can be seen as embracing some hybrid version of theory theory and simulation theory, where "theory theory" (and appeal to folk psychology) is a provisional conception that requires neurobiological refinement (Churchland 1986: 299ff; 2003: 107, 111ff). Whether this refinement will eventually lead to a consensus around the idea of simulation – which, as Churchland (2003: 108) notes, is a move that looks extremely promising from the perspective of recent neuroscience of resonance systems and mirror neurons (and she is not alone in this – see, e.g., Gallese and Goldman 1998; Jeannerod and Pacherie 2004; Metzinger 2003: 368ff) – is still an open question. Rebecca Saxe (2005), for example, using evidence from experimental and developmental psychology and neuroscience, argues against simulation theory and the idea that it is instantiated at a neural level in mirror neurons. But the only alternative that she considers is theory theory, and she cites neurological evidence in support of the latter.

Recent phenomenological critiques of both theory theory (TT) and simulation theory (ST) challenge these neurophilosophical readings and offer an alternative that is not unrelated to Ryle's view. Since I've offered a critique of TT on both phenomenological and scientific grounds elsewhere (Gallagher 2001, 2004), and since ST seems to be the ascending theory these days, notwithstanding Saxe's analysis, let me set out the phenomenological critique of simulation.

---

[4] Actually, Dreyfus notes the connection between Heidegger and Ryle, and what Ryle could have learned from Heidegger: "Heidegger offers a phenomenological analysis of everyday masterful, practical know-how that dispenses altogether with the need for mental states like desiring, believing, following a rule, and so on, and thus with their intentional content" (Dreyfus 1987).

Arguments againsts explicit ST: First, on the traditional view of ST, as embraced by Goldman (1989) for example, simulation is explicit, that is, at least a partly conscious, introspective strategy. "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 2005). In fact, Quine expressed this view in 1960, perhaps at that meeting in Paris, but certainly in his book *Word and Object*.

> Casting our real selves … in unreal roles …. we find ourselves attributing beliefs, wishes and strivings even to creatures lacking the power of speech, such is our dramatic virtuosity. We project ourselves even into what from his behaviour we imagine a mouse's state of mind to have been, and dramatize it as a belief, wish or striving … (Quine 1960: 219).

If this concept of simulation is proposed as the primary and pervasive way that we gain understanding of others, then phenomenology offers several objections. The first is the one mentioned by Ryle: "the 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" (1960: 54). This is similar to an objection raised by Scheler (1954) against the forerunner of ST, the argument (inference) from analogy. Scheler offered another objection: that this kind of inferential simulation is too cognitively complex to account for the infant's ability to understand the intentions of others (for which Meltzoff 1995 provides evidence in 18-month-old children). Finally, by straight appeal to phenomenology (that is, by consulting our experience) we can develop what I call the "simple phenomenological objection." Namely, it seems clear that running conscious simulation routines is not a pervasive way that we attempt to understand others. This does not mean that we never use this kind of simulation, but reflection on our own experience should tell us that the use of this kind of simulation is relatively specialized and rare.

Arguments against implicit ST: Second, if, as in more recent formulations of ST, the claim is that the simulation is subpersonal, instantiated in the workings of mirror neurons, shared representations, or resonance systems, and is therefore automatic and nonconscious, then it seems that straight up phenomenology can make no objection since its scope is limited to conscious processes. Here, however, we can appeal to neurophenomenology and ask our question again: What precisely are we trying to explain? Are we explaining simulation (and what precisely does that mean) or are we explaining perception? In response, let me make two arguments against implicit (subpersonal) simulation theories.

First argument against implicit ST: implicit ST operates with the wrong concept of perception. Mirror neurons fire 30-100 ms after appropriate visual stimulation (e.g., seeing another person perform a simple intentional act) (Gallese, personal communication). The perception of another person performing a simple intentional act is itself something that requires time (both the formation of the act and the perception). But let's say that at some moment the other person's movement registers in the activation of the visual system, and 30-100 ms after that we get activation of the mirror system. Now the question is precisely were to draw the line between the act of perception and the simulation. At least on traditional versions of ST, simulation is identified as a mental process distinct from perception. It is offered as a step-wise process that begins with perception and ends with inference. We first see an action. Then we simulate it in our own mind or motor system. Then we infer something about the other's experience. Does this articulation remain convincing in the 30-100 ms of sub-personal processing that is described as the automatic mirroring simulation?

The phenomenological alternative here is that in most cases, when I see the other's action or gesture, I *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. This is consistent with the idea of enactive perception as outlined by Varela, Thompson, Noë, O'Regan, and others. The act of perception is defined not

simply as a sensory activation; it includes motor components. So the line between neuronal activation in the visual system and neuronal activation of the mirror system is not a line that we can draw between perception and simulation; the resonating effect is part of the perceptual process. In that case, mirror activation is not the initiation of simulation; it's part of a direct perception of what the other is doing. Although Jeannerod and Pacherie (2004) defend a version of ST, they nicely express the phenomenological alternative: "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).

Second argument against implicit ST: What implicit ST calls 'simulation' is not really simulation. What do we mean by simulation? Traditionally, a simulation is an instrumental *model* we *use* to try something out when we can't work with the real thing. In the psychological and philosophical literature, ST contends that rather than using a theory about what other people believe or intend, we *use* our own mind or our own motor system as a model to simulate (we imagine, or we try out "*pretend* beliefs" about) what's going on in their mind (their beliefs, desires, etc.) in order to understand their actions. Two things seem essential to this ST concept of simulation. I'll call them simply, the instrumental aspect, and the pretense aspect.

We can find both aspects in the literature of ST. Consider the following characterizations (italics are mine).

- *Instrumenal aspect:* "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)

- *Instrumental aspect:* "Simulating others -- i.e., *using* one's own evaluation and reasoning mechanisms as a model for theirs …" (Dokic and

Proust 2002).

- *Instrumental aspect:* "*Using* our own motor capacities to understand the actions performed by others is at the core of the simulation theory. … the neural motor system involved in the preparation and execution of action, is also part of a simulation network which is *used* to interpret the perceived actions performed by others" (Chaminade, Meary, Orliaguet, Decety 2001)

- *Pretense aspect:* [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. 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,

- *Pretense aspect:* "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).

Simulation, then, 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 said to be "deliberately adopted"), and it involves a pretense

condition (I put myself "as if" in the other person's shoes).

But I think we are justified in asking: In what sense is the subpersonal activation of the mirror system or shared representations still simulation as ST defines it? The activation of mirror neurons and shared representations are *not* models that we create or initiate or use -- they are *reactions* that are initiated by others. The intentional actions of others *elicit* this resonance activation, and by so doing, such resonance processes directly deliver a sense of what the other intends. If we define simulation as a model that we use, these processes are not simulations. Rather, the other person's action that we directly observe (and we rightly might ask why we need a model when we are looking at the real thing?) automatically activates in our brain the same areas that are activated when we act in a similar manner. The other person has an effect on us by eliciting this activation. On that basis we get a good sense of what their intention is. This is not a simulation, but a perceptual elicitation. It's not us doing it, but the other who does this to us.

Furthermore, there is no neuronal subjunctive, that is, the pretense condition is not met, because at the neuronal level there is no "*as if* it were I," or "*as if* I were you" involved, and in that regard, too, it fails to be the kind of simulation required by ST. 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. A 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.

Concluding scientific prescript

Ryle suggested that philosophy of mind needed neither phenomenology nor empirical science. Contemporary neurophilosophers are quite willing to ignore phenomenology and place all their bets on science. Neurophenomenologists, from Merleau-Ponty to Varela (see Varela, Thompson, and Rosch 1991; Thompson and Varela 2001), have prescribed that we take both phenomenology and empirical science quite seriously.

Is it possible to be both phenomenological and scientific about consciousness, intersubjectivity, and other topics that are of interest in philosophy of mind? In various contexts I keep running into the question: What does it mean to be scientific? People often appeal to the idea of the third-person objectivity of scientific procedure to answer this question. Some people think that science is restricted to quantitative accounts, and that if something cannot be quantified, it doesn't allow for scientific study. In neurophilosophy and the cognitive sciences there are people who will insist that the task of science is to be reductionistic: a good account is one that can be mapped out completely in sub-personal terms.

I think that it is better to think of science as using any means possible to explain *what there is*. And if *what there is* includes such things that cannot be reduced to computational processes or the subpersonal activation of neurons, or cannot be quantified, or objectified without loss – such things as mentioned by Ayer, "the whole ensemble of processes, of manners of being, of actions, of sensations, or of impressions that one cannot consider as objects," processes and things that

nonetheless have meaning for human life, and that therefore fall into the province of phenomenology, hermeneutics, and the humanities – then to turn away from them and to deny their actuality is in fact being unscientific.

References

Ayer, A. J. 1960. Phenomenology and analytic philosophy. In Merleau-Ponty (1992).

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

Carruthers, P. 1998. Natural theories of consciousness. *European Journal of Philosophy* 6/2: 203–222.

Chaminade, T., Meary, D., Orliaguet, J-P., Decety, J. 2001 Is perceptual anticipation a motor simulation? A PET study. *Neuroreport* 12 (17) 4.

Churchland, P. S. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P.S. 2002. *Brain-Wise: Studies in Neuro-philosophy*. Cambridge, MA: MIT Press.

Dennett, D. 1996. Tiptoeing past the Covered Wagons in "Dennett and Carr Further Explained: an exchange" *Emory Cognition Project*, Report #28, Department of Psychology, Emory University, April 1994. http://ase.tufts.edu/cogstud/papers/tiptoe.htm

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

Dreyfus, H. 1987. In Magee, B. 1987. *The great philosophers*. Oxford: Oxford University Press.

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. 2003. Phenomenology and experimental design. *Journal of Consciousness Studies* 10 (9-10): 85-99.

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

Gallagher, S. and Jesper Brøsted Sørensen. (2006). Experimenting with phenomenology. *Consciousness and Cognition* 15 (1): 119-134

Gallagher, S. and Zahavi, D. 2005. Phenomenological approaches to self-consciousness. *Stanford Encyclopedia of Philosophy* (at http://plato.stanford.edu/).

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

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

Goldman, A. 2005. Mirror systems, social understanding and social cognition. *Interdisciplines* (www.interdisciplines.org/mirror/papers).

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. 1989. Interpretation psychologized. *Mind and Language* 4: 161-85.

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

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.

Jeannerod, M. and Pacherie, E. 2004. Agency, simulation and self-identification. *Mind and Language* 19 (2): 113-146.

Merleau-Ponty, M. 1992. *Texts and Dialogues*. H. J. Silverman and J. Barry (eds.). Atlantic Highlands, NJ: Humanities Press.

Metzinger, T. 2003. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Meltzoff, A. N. 1995. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology,* 31: 838-50.

Quine, W. 1960. *Word and Object*. Cambridge, MA: The Technology Press of M.I.T; New York & London: Wiley, 1960.

Rosenthal, D. 1993. Higher-order thoughts and the appendage theory of consciousness. *Philosophical Psychology* 6: 155-66.

Ryle, G. 1960. Phenomenology and analytic philosophy. In Merleau-Ponty (1992).

Ryle, G. 1949. *The Concept of Mind*. New York: Barnes Noble.

Saxe, R. 2005. Against simulation: The argument from error. *Trends in Cognitive Sciences* 9 (4): 174-79.

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.

Thomasson, A. L. 2002. Phenomenology and the Development of Analytic Philosophy. *Southern Journal of Philosophy* 40, Supplement: 115-142.

Thompson, E. and Varela, F. 2001. Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences* 5 (10): 418-25.

Varela, F., Thompson, E. & Rosch, E. 1991. *The Embodied Mind.* Cambridge, MA: MIT Press.

Zahavi, D. 2002. First-person thoughts and embodied self-awareness: Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences* 1: 7-26.

Zahavi, D. and Parnas, J. 1999. Phenomenal consciousness and self-awareness: A phenomenological critique of representational theory. In S. Gallagher and J. Shear (eds.), *Models of the Self* (pp. 253-70). Exeter: Imprint Academic.