

This preprint version is posted here by permission of Taylor & Francis for personal use, not for redistribution. Please cite the definitive version, which was published in *Philosophical Explorations*, 11(3), 223-235. It can be downloaded at: <http://dx.doi.org/10.1080/13869790802239268>

## **Folk Psychological and Phenomenological Accounts of Social Perception**

Mitchell Herschbach

Address:

Department of Philosophy  
University of California, San Diego  
9500 Gilman Drive #0119  
La Jolla, CA 92093-0119

Email: [mherschb@ucsd.edu](mailto:mherschb@ucsd.edu)

Website: <http://sdcc3.ucsd.edu/~mherschb/>

Abstract:

Theory theory and simulation theory share the assumption that mental states are unobservable, and that mental state attribution requires an extra psychological step beyond perception. Phenomenologists deny this, contending that we can directly perceive people's mental states. Here I evaluate objections to theory theory and simulation theory as accounts of everyday social perception offered by Dan Zahavi and Shaun Gallagher. I agree that their phenomenological claims have bite at the personal level, distinguishing direct social perception from conscious theorizing and simulation. Their appeals to phenomenology and other arguments do not, however, rule out theory theory or simulation theory as accounts of the subpersonal processes underlying social perception. While I here remain uncommitted about the plausibility of subpersonal theorizing and simulation, I argue that phenomenologists must do more in order to reject these accounts.

Keywords: folk psychology; phenomenology; simulation theory; social perception; theory theory

### **1. Introduction**

The standard account of human social understanding is that we understand people's behavior in terms of their mental states. This "folk psychological" account comes in two major versions: theory theory and simulation theory. A core assumption of both theory theory and simulation

theory is that we only perceive or experience the physical movements of persons, since mental states are unobservable. The two views differ with regard to the kind of psychological process used to go from observable behavior to mental state attributions. Mental state attribution occurs for theory theory via theoretical inference, by applying theoretical knowledge about the relations between observable behavior, environmental context, and mental states. Simulation theory denies that we possess such theoretical knowledge, instead claiming that we simulate being another person, determine what mental states we would have if we were that person in that situation, and project those mental states onto the other person. For both folk psychological accounts, our perceptual experience is of mere behavior, and mental state understanding only comes when we perform some extra psychological process beyond perception.

This folk psychological picture of social perception has recently come under attack by philosophers appealing to phenomenological considerations (Gallagher 2001, 2005, 2007, forthcoming; Gallagher and Zahavi 2008; Ratcliffe 2006; Zahavi 2005, 2007). They instead claim that we perceive people's emotions, intentions, etc., in their behavior, that behavior is expressive of mentality and not divorced of all psychological or mental import. Zahavi, for example, writes that "...the life of the mind of others is visible in their expressive behavior and meaningful action" (2005, 222). Scheler and Wittgenstein are often identified as historical proponents of this view, with passages such as the following:

For we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of

his thoughts in the sound of his words. If anyone tells me that this is not ‘perception’, for it cannot be so, in view of the fact that a perception is simply a ‘complex of physical sensations’, and that there is certainly no sensation of another person’s mind nor any stimulus from such a source, I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts. (Scheler 1954, 260)

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

Authors differ in their descriptions of this experience of other persons, calling it either a form of “direct perception” (Gallagher 2007, forthcoming) or a “distinctive mode of consciousness, different from perception, recollection and fantasy” called “empathy” (Zahavi 2007, 36). But all those in this camp agree that we can “directly”—i.e., without a mediating conscious psychological process, such as inference—experience other people’s mental lives. Further, they claim that direct social perception is much more pervasive in our everyday social lives than the conscious reflection emphasized in folk psychological accounts.

As Gallagher (forthcoming) notes, the idea of “direct perception” is often associated with Gibson (1979). By “direct perception,” Gibson meant that all the information we need to perceive objects or possibilities for action is already in the light transmitted to our sense organs, ready to be “directly” detected by us; it need not be inferred or computed from sensory

stimulation via additional psychological processes. What exactly Gibson meant by “directly” is a matter of dispute. One interpretation mirrors the phenomenologists’ sense of direct perception, reading “no additional psychological processes” as meaning no additional *conscious* psychological processes. But even granting this, no one can deny that perception involves complex processes inside a person, particularly their brain. In other words, even given the directness of perception at the *personal level* (i.e., the level of phenomenological experience), there remains a story to be told at the *subpersonal level* of the brain processes (and potentially other non-neural internal processes) enabling perception. Gibson is also often read as rejecting information-processing or representational characterizations of these subpersonal processes. Whatever Gibson’s views on the subject, it is a central issue for cognitive science to characterize the subpersonal level processes enabling personal level psychological phenomena such as perception.

Focusing on this distinction between personal and subpersonal level accounts, I will here evaluate Dan Zahavi and Shaun Gallagher’s respective objections to theory theory and simulation theory as accounts of social perception—i.e., the experience we have of other people when perceiving their behavior. I will argue that their phenomenology-based criticisms are much more narrowly focused than they appear. While they do have bite against *personal level* versions of theory theory and simulation theory, they do not rule out these theories as accounts of the *subpersonal level* processes enabling direct social perception. Further, I show that their arguments directly addressing the subpersonal level (particularly Gallagher’s rejection of a subpersonal notion of simulation) are unconvincing. To be clear, in this paper I remain agnostic about theory theory and simulation theory as accounts of the subpersonal processes underlying direct social perception. Indeed, I leave open whether *any* subpersonal processes should be

characterized in these terms. My aim is instead to distinguish personal and subpersonal level accounts of direct social perception and expose the limitations of Zahavi and Gallagher's criticisms about subpersonal level versions of theory theory and simulation theory.

## **2. Zahavi on theory theory**

In his book *Subjectivity and Selfhood*, Dan Zahavi (2005, chap. 7) presents arguments against theory theory's account of self- and other-awareness. Although the two are related, I will focus on Zahavi's arguments against theory theory's account of other-awareness—specifically, our ability to attribute mental states to others, what is often called “mindreading.” Zahavi's challenge to theory theory focuses on the connections it draws between mindreading, possessing a theory of mind, and passing false belief tasks. It can be summarized as follows (2005, 197, 214):

1. Theory theory proposes that to mindread, one must possess a theory of mind.
2. According to theory theory, passing false belief tasks is necessary and sufficient for possessing a theory of mind.
3. Children do not pass false belief tasks until around age four.
4. Therefore, according to theory theory, children do not possess a theory of mind until around age four.
5. Therefore, theory theory is committed to the claim that children cannot understand others' mental states until around age four.

6. But children can perceive others' emotions and intentions prior to age four (for evidence from developmental psychology, see Zahavi 2005, 206–14).
7. Thus, theory theorists must either: (a) adopt an inclusive definition of mindreading and admit that children can mindread before they can pass theory of mind tasks (i.e., false belief tasks); or (b) retain an exclusive definition of mindreading in terms of false belief understanding, but concede that children acquire an understanding of emotions and intentions prior to being able to mindread (understood as theorizing about the mind).

The force of the dilemma in (7) is supposed to be that, either way, theory theorists must admit that at least some mental states are understood non-theoretically. Unfortunately, several aspects of this argument are problematic. I will not address evidence against (3), such as Onishi and Baillargeon's (2005) purported evidence of false belief understanding in 15-month-olds, which potentially collapses the developmental gap between false belief understanding and the understanding of emotions and intentions. What I will focus on is premise (2), and its role in the dilemma Zahavi poses in (7). Most researchers, including theory theorists, now accept that undue attention has been paid to false belief understanding (see, e.g., Bloom and German 2000).

Accordingly, most theory theorists now reject (2). While false belief understanding is sufficient for possessing a fairly mature theory of mind, it is no longer treated as necessary; understanding of mental states besides belief is treated as evidence for possessing a theory of mind. In other words, theory of mind is now treated as multifaceted, with some aspects developing prior to false belief understanding.

Given its dependence on (2), the dilemma in (7) is thus problematic. While the two horns differ in what falls under the heading of mindreading, both depend on (2), which identifies tests

of theory of mind with tests of false belief understanding. As noted above, however, theory theorists no longer treat false belief understanding as criterial for possessing a theory of mind. Therefore, the force of Zahavi's dilemma for theory theory is undermined.

There is, however, a substantive issue that emerges from Zahavi's discussion, one which he himself takes up on the last pages of his book (2005, 221–2). Zahavi admits that we do sometimes consciously theorize about people's mental states, and, following Frith (2003), suggests that this is how some high-functioning autistic people are able to overcome some of their deficits in social understanding. The evidence Zahavi provides in favor of (6)—about children's understanding of others' emotions and intentions—does not, however, involve such conscious, reflective understanding. For example, Zahavi (2005, 212) describes the capacity for social referencing found in children during their second year of life. When in an unfamiliar situation, infants look to their parent's faces to gauge their emotional reactions to the situation. Infants then use this information about the status of the environment to guide their own actions. If a parent expresses, e.g., fear toward an unfamiliar object, the child will recognize this negative reaction toward the object and, accordingly, avoid it. It is unreasonable to characterize infants as consciously reasoning from facial expressions to emotional reactions; further, it does not fit the phenomenological experience of adults, as characterized by the quotations from Scheler and Wittgenstein above. As a result, Zahavi describes such instances of social perception as “immediate, pre-reflective, or implicit understanding” of others' mental states (2005, 221).

Therefore, we can phenomenologically distinguish two capacities: (a) the ability to consciously theorize about people's mental states; and (b) the ability to directly, i.e., non-inferentially, perceive their mental states. Allowing that theory theory may explain the former, the key question then concerns whether our capacity for phenomenologically direct social

perception is driven by *unconscious, subpersonal* theorizing. Zahavi's clearly thinks it is not. But what is the evidence for this conclusion?

Unfortunately Zahavi does not offer much of a defense. He describes Baron-Cohen's (1995) proposed "intentionality detector" and "eye-direction detector" as providing "direct and non-theoretical understanding" of people's intentions and perceptions (Zahavi 2005, 214). But Zahavi offers nothing to support the claim that such mechanisms are in fact non-theoretical. This also comes up in Zahavi's discussion of autism, mentioned briefly above. Uta Frith (2003) claims that autistics lack a "theory of mind module that allows for an intuitive and automatic attribution of mental states to others," but that autistics "might acquire a conscious theory of mind by way of compensation" for this lack (Zahavi 2005, 222). Zahavi accepts that autistics may consciously and explicitly theorize about people in order to make up for their social deficits, but rejects Frith's characterization of nonautistic social perception as theoretical. Yet he does no more than assert this claim, writing that it is "better to avoid using the term 'theory' when speaking of a nonconscious information-processing mechanism" involved in social perception, and that he finds "it rather misleading to designate such nonconscious inferential processes as intuitive" (2005, 222).

I do not want to defend the idea that autistics' compensatory, conscious theorizing about people's minds should be equated with nonautistic social perception. Surely there are differences between these phenomena. My point is that the phenomenological considerations offered by Zahavi are insufficient to rule out theory theory as an account of ordinary social perception. Phenomenological claims simply do not have purchase with regard to nonconscious, subpersonal processes, as Gallagher (2005, 215) admits. Theory theorists do usually characterize our social experience as of mere behavior, from which we infer people's mental states. Yet as a claim about

subpersonal processes, it seems theory theory could accommodate the idea that our personal level, phenomenological experience is not of mere behavior, but of “expressive behavior,” i.e., behavior expressive of people’s mental states (Zahavi 2007). Perhaps theory theory needs an enriched account of our mental state concepts to accommodate this fact. It looks plausible, however, that the subpersonal processes producing such experience might be theoretical in nature. Phenomenological evidence will not speak to this possibility.

So what is it that grounds Zahavi’s rejection of theory theory at the subpersonal level? At the beginning of his chapter on this topic, Zahavi (2005, 181) mentions Blackburn’s (1992) “promiscuity objection” to theory theory: the concern that characterizing subpersonal processes as “theoretical” would make this concept entirely vacuous, so almost any belief-formation process would count as theoretical. Zahavi seems persuaded by this worry, such that he treats theorizing and the related concepts of explanation and prediction as exclusively personal level concepts characterizing conscious, reflective phenomena. Gallagher explicitly asserts this claim, writing: “Explanation (or theory) seems to mean (even in our everyday psychology) a process that involves reflective consciousness” (2005, 215). Such a strict conception of theorizing rules out the possibility that (fully) subpersonal processes—such as those enabling direct social perception—are theoretical in nature.<sup>1</sup>

I acknowledge Zahavi’s concern with trivializing the concept of theorizing. Proponents and even critics of theory theory have often been quite permissive in the use of this concept. For example, Shaun Nichols and Stephen Stich use “theory theory” to refer to any “information-rich” process of mental state attribution, i.e., any process which is guided by “a rich set of mental representations containing substantial information (or, sometimes, *mis*information) about mental states and their interactions with environmental stimuli, with behavior, and with each other”

(2003, 102). Simulation processes, which do not require such bodies of information, are cases of “information-poor” processes. Such a characterization of theory theory would indeed allow many different kinds of subpersonal psychological processes to be characterized as theoretical. For example, this definition of theory theory does not mark the distinction between “classical” cognitive architectures (i.e., “rules and symbols” approaches sometimes derided with the label “good old fashioned artificial intelligence”) and connectionist architectures. Some have thought true theorizing to require the rule-based manipulation of symbolic representations, and thus rejected connectionist networks as vehicles of theorizing. Moving from features of representational vehicles to issues of representational content, this permissive characterization of theory theory covers any information about mental states, environmental conditions, and behavior. It does not matter whether this information is a unified, coherent, abstract set of laws, or a less cohesive collection of statistical patterns, algorithms and heuristics. It has been argued that only the former and not the latter deserve to be called “theoretical” in nature.

But even those who use the term “theory” rather permissively admit such distinctions between types of subpersonal processes and draw stronger or weaker connections between these various phenomena and the conscious theorizing of adults (scientists in particular). Gopnik and Meltzoff (1997), for example, quite specifically define theories in terms of their structural features (abstract, coherent representations of the causal structure of some target domain), functional features (their use for explanation, prediction, and interpretation) and dynamic features (their defeasibility in light of counterevidence, and the nature of intertheoretic change). Notably, unlike Gallagher and apparently Zahavi, they do not make consciousness a criterion for theory possession, formation, use, or change. With such a specific conception of theories, Gopnik and Meltzoff rightly admit that not all of our knowledge is theoretical, and identify other

types of cognitive mechanisms they call “modules” and “empirical generalizations.” They characterize modules as possessing the same structural and functional features as theories, but unlike theories, as being resistant to counterevidence. Empirical generalizations, which include scripts, schemas, and narratives, are knowledge structures more tied to immediate experience. Because of this, they lack the abstractness and coherence of theories, and thus differ in their explanatory and predictive capacities. Gopnik and Meltzoff loosely follow Fodor (1983) on how these mechanisms are interrelated, with modules serving as input systems to “central” cognition, the place of theories and empirical generalizations. Let me be clear that I am not advocating Gopnik and Meltzoff’s taxonomy of cognitive mechanisms, or the connections they draw between children’s knowledge and scientific theorizing. I mention their account because it is illustrative of how theory theorists could meet the promiscuity objection without requiring theorizing to be a personal level, conscious, reflective process. Subpersonal brain mechanisms might be characterized as possessing representations, developing theories, etc., without their requiring personal level forms of consciousness and intelligence.

Thus, if Zahavi wants to reject a subpersonal conception of theorizing, so social perception cannot be accounted for by theory theory, he must be more specific about what features of personal level theorizing are objectionable at the subpersonal level. For example, does Zahavi understand “theorizing” as making inferences with propositionally structured representations, as in the deductive-nomological model of explanation once so dominant in the philosophy of science? Or is theorizing intended more broadly, to cover the connectionist-style processing Paul Churchland (1989, chap. 9) believes to characterize the brain’s cognitive operations? Are both of these problematic forms of “theorizing”? Zahavi refers to subpersonal “*information-processing mechanisms*” (2005, 222, italics added), and thus might be open to

representationalist subpersonal accounts of social perception. But nothing further is said about the nature of such subpersonal representations, so perhaps even this interpretation of Zahavi's position is too quick. A nonrepresentational alternative might be found in the account of emotion perception recently argued for by Dan Hutto, and endorsed by Gallagher (in Menary 2006; see also Hutto 2008). Note, however, that in making this argument, Hutto directly addresses criteria for ascribing informational or representational content to subpersonal mechanisms. This is a very different kind of argument from Zahavi's phenomenology-based criticism of the pervasiveness of conscious theorizing, which is pitched at the personal level. Zahavi would similarly need to argue in a way appropriate to the subpersonal level in order to rule out subpersonal theorizing as underlying direct social perception.

As this cursory discussion shows, there are a number of interesting issues for philosophers and psychologists about the subpersonal processes underlying social perception. While phenomenological evidence is important in providing adequate personal level accounts, it alone is insufficient to rule out subpersonal theorizing as enabling social perception. Different kinds of evidence are needed to evaluate claims about subpersonal processes. In this case, an extended discussion of possible subpersonal explanations is required if Zahavi is to adequately evaluate theory theory as an account of social perception.<sup>2</sup> While I am not here endorsing theory theory as an adequate characterization of the subpersonal processes enabling direct social perception (or, indeed, an adequate description of *any* subpersonal processes), my point is that Zahavi's arguments have failed to rule theory theory out.

### **3. Gallagher on simulation theory**

A similar dialectic occurs in Shaun Gallagher's (2001, 2005, 2007, forthcoming) recent criticisms of simulation theory (I will focus mainly on the arguments in Gallagher 2007). Gallagher objects to conscious simulation as an account of our everyday social experience by appeal to phenomenology. His "*simple* phenomenological argument" (2007, 356) is that we just do not very often find ourselves consciously simulating others' mental states. This parallels Zahavi's denial that conscious theorizing pervades our everyday social interactions. Gallagher does not deny that we sometimes engage in conscious simulation, but claims that this is relatively rare and thus cannot account for how we understand all the people we perceive in our daily lives. Gallagher (2007) also objects to the claim that conscious awareness of simulation diminishes as it becomes habitual, analogous to the way our driving habits recede out of conscious awareness as we become expert drivers. This is because, Gallagher claims, even habitual processes can become objects of conscious reflection. And we do not seem capable of turning cases of social perception into ones of explicit simulation through reflection. Social perception should thus be seen as a distinct phenomenon from conscious simulation.

These objections to simulation theory seem right as personal level claims. Conscious simulation is phenomenologically distinct from social perception. This provides an important corrective to claims about the pervasiveness of conscious simulation. If it is to account for more than the relatively rare cases of conscious imaginative simulation, simulation theory must be treated as a theory of the subpersonal processes underlying direct social perception. As Gallagher (2007) recognizes, phenomenological evidence will not speak to the nature of these subpersonal processes. To reject this version of simulation theory, Gallagher thus makes a conceptual argument, contending that subpersonal processes do not meet simulation theory's own criteria for something to be a simulation.

Gallagher begins with two definitions of ‘simulation’ offered by the Oxford English Dictionary: (a) the *pretense definition*: “Simulation is an imitation, in the sense of something *not real*—counterfeit; to simulate means to feign, to pretend”; and (b) the *instrumental definition*: “a simulator: a model (a thing) that we can use or do things with so we can understand the real thing” (2007, 359). Quoting descriptions of simulation theory by various authors, Gallagher argues that simulation theory combines these two definitions, such that simulation is a process where I use (i.e., “control in an instrumental way”) my own psychological mechanisms as pretend or “as if” models of another person (2007, 360). Gallagher then rejects the notion of subpersonal simulation by arguing that neither the instrumental nor the pretense condition is present at the subpersonal level.

First, with regard to the instrumental condition, if simulation theory requires that “I (or my brain) uses or controls” a simulation, Gallagher denies that such control occurs at the subpersonal level (2007, 360). The core of Gallagher’s objection is that the instrumental or control condition is best understood at the personal level. If simulation is a reflective, personal level process, it is at least partly under conscious control—something we can initiate and terminate at will. But at the personal level we do not control our subpersonal processes in this way, and thus do not “use” implicit simulations to model the other person’s mental states. Going fully subpersonal, Gallagher rejects the claim that the *brain* uses any neural processes as simulations. The objection here appears to be largely conceptual, claiming that the personal level concepts of use and control do not make sense at the subpersonal level. Rather than simulation being a controlled process following perception, Gallagher argues that the neural processes activated when we perceive a person are “elicited” in us. Hence, mirror neuron activation or other kinds of neural processes should not be described as simulations, but rather as part of the

temporally extended process of social perception.

I'm on board with Gallagher that subpersonal processes as such are not subject to personal level control. But it seems overly restrictive to say that one neural process is not "used" by another. Of course one commits the homuncular fallacy to literally apply the personal level concept of "use" to brain mechanisms. Brain mechanisms do not use other brain mechanisms as representations or models in the same way persons understand and use external representations and models (e.g., maps, texts, scale physical models, etc.). But to imply that this is the only acceptable sense of "use" or "control" begs the question against a subpersonal version of simulation theory, and the idea of subpersonal representation in general.

A defense of the instrumental condition for simulation theory can be found in William Ramsey's (2007) recent critical evaluation of the appeal to representations in cognitive science. Ramsey defends precisely the subpersonal notion of using models as representations rejected by Gallagher. Like simulation theorists, Ramsey argues that such models serve as representations—i.e., are able to "stand in" for the things in the world—by being structurally similar or isomorphic to the things they represent.<sup>3</sup> My focus here, though, is on Ramsey's discussion of how subpersonal mechanisms can be "used" by other subpersonal mechanisms without requiring those mechanisms to be intelligent homunculi. Ramsey (2007, 194–203) makes his case using the example of a car navigating its way along an S-shaped track. If a real person were driving the car, one way they could steer the car would be to use a map of the track. This is an obvious case of using a model. Ramsey then considers removing the driver, turning the car into a mindless system. Could the internal workings of the car be automated such that they could still be characterized in terms of using a model of the track?

One way we might do this, suggested by Cummins (1996), would be to convert the S-curve of the map into an S-shaped groove into which a rudder would fit. The rudder could then move along the groove as the vehicle moves forward, and the direction of the steering wheel and, thus, the vehicle's front wheels could be made to correspond to the direction of the rudder.... As the rudder moves along the groove, its change in orientation would bring about a change in the orientation of the front wheels. Because the shape of the groove is isomorphic with the curve itself, the wheels change along with the S-curve and the vehicle moves through it without ever bumping into a wall. (Ramsey 2007, 198)

The car's internal workings could be characterized without any appeal to representations, as can be done (in principle) with any representational system. But the most natural way of explaining how the car navigates the road is that the groove serves as a map of the course of the track, with sections of the groove "standing in for" segments of the track. The sense in which the groove is being "used" as a model by the other parts of the car does not require any intelligence on their part. As Ramsey summarizes, "A mindless system can still take advantage of the structural isomorphism between internal structures and the world, and in so doing, employ elements of those internal structures as representations-qua-stand-ins" (2007, 200).

In the same way, brain mechanisms could "use" other brain mechanisms as models without requiring the intelligence of a person. That the task of social perception is one of understanding rather than of pure behavioral navigation should not be a barrier to characterizing the subpersonal processes in terms of simulation, of using a model. In addition, whether such a process is initiated endogenously or activated in response to stimulation from the environment does not seem relevant to whether such a process should be described as "using a model." This

cuts against Gallagher's characterization of the neural processes underlying social perception not as simulations but as "effects" which are "elicited" in us by the other person's presence (2007, 360–1).

Gallagher seems to recognize this possibility of subpersonal processes meeting the instrumental condition when discussing accounts of social perception based on the motor control literature (e.g., Hurley 2005, 2006), which apply just such an understanding of neural processes "using" other neural processes as models.<sup>4</sup> Gallagher does not mention any problem with the instrumental condition for these accounts, and instead argues that they fail simulation theory's pretense condition. The core idea of these accounts is that when motor commands are sent to control the body, efference copies of these motor commands are sent to a mechanism called a "forward model" to predict the success of these movements toward achieving one's goals. One useful feature of these predictions is that they can drive corrections to behavior faster than can be done by sensory feedback. This picture of motor control is then used to offer a simulation-based account of action perception: the forward model or models used in motor control to predict one's own movements are co-opted in social perception to simulate the actions of others and allow predictions of their behavior, and through additional processes, the mental states causing these actions. As Gallagher summarizes, these motor-based accounts claim that "the perception of the other's action is automatically informed by a sub-personal simulation; perception of action involves a loop through the [forward model]" (2007, 362, n. 10).

Given its uncontroversial sense of "using a model," Gallagher objects that the motor representations in these accounts fail to meet simulation theory's pretense condition. Expressing this objection, Gallagher writes that "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'" and that

implicit simulation requires a “representation of my own motor action *as if it were* the other’s” (2007, 362, n. 10). As I read him, Gallagher believes that the pretense or “as if it were I”<sup>5</sup> component must be in the content of a representation for it to count as a subpersonal simulation. It is not enough if these states are explicitly marked as belonging to another person rather than oneself. For Gallagher, to count as simulations, they must be represented as states of oneself “pretending” to be those of another person. As he writes about mirror neurons:

...the mirror system is neutral with respect to the agent; there is no first- or third-person specification involved. In that case, they do not register *my* intentions as pretending to be *your* intentions; there is no “as if”—there is no neuronal subjunctive—because there is no “I” or “you” represented. (2007, 361)

The “motor simulations” of motor-based accounts do not have content of the kind Gallagher requires. These motor representations activated during social perception are simply the same ones activated during motor control. As normally characterized, they do not explicitly mark the agent in question (self or perceived other), so they certainly cannot have as part of their content that this is the motor representation I would have if I were the other person. In this way Gallagher argues that motor-based accounts of social perception fail to involve subpersonal simulation.<sup>6</sup>

The pretense condition clearly applies to personal level simulation where I consciously imagine being another person. But we might question Gallagher’s reading of simulation theory’s pretense condition that it requires any state/process counting as a simulation to include an “as if it were I” component in its representational content.<sup>7</sup> Gallagher defends this reading by writing:

For [simulation theory], a simulation is not simply a model that we use to understand the other person—theoretical models would suffice if this were all that was 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. (2007, 360)

Of course simulations must be distinguished from theoretical models. This can be done by distinguishing the kinds of representations involved. Folk psychological theorizing requires representations of folk psychological generalizations or laws: how mental states relate to environmental stimuli, behavior, and each other. It is precisely such representations which simulation theory denies.<sup>8</sup> Simulation theory instead posits mental states/processes which *replicate or resemble* those of the target system being simulated. These are often characterized as “pretend” mental states, but this is not necessary: the essential feature of simulation is that the simulation replicates or is similar to the target state being simulated (Goldman 2006). Of course this simulated process concerning another person must be distinguished from a genuine process concerning myself, and be used to create explicit representations of the mental state or states as the other person’s.<sup>9</sup> This is a problem for all so-called “shared representations,” i.e., agent-neutral representations of the properties of self and other. But all simulated states need not be represented as being “as if” they were another person’s. Aspects of a simulation process may go on without the identity of the agent (self vs. other) being explicitly represented. Just because the end result of this process must be categorized and attributed to an agent does not mean all aspects of the purported simulation process must explicitly represent the agent in question to count as simulations. To adapt some of John Perry’s (1993, chap. 10) terminology, they may

*concern* the simulated person without being *about* (i.e., explicitly representing) that person during the simulation process.

Imagine that perception of another person's, say, anger occurs by activating (a) motor representations associated with the facial expression tokened by that person and (b) other neural states associated with the experience of anger (as Goldman 2006 suggests is actually the case). Following many characterizations of simulation theory, it makes sense to say that the "online" function of these states is for my own experiences of anger, and that they are being used "offline" for understanding another person's anger. Because of this, these "shared representations" would arguably replicate or resemble those of the perceived angry person. While these states do not *represent* "the motor/emotional state I would be in if I were that other person," they are the states *I would be in if I were myself angry*. While such states must still be used to create a distinct representation *that he is angry*, the fact that I would be in these states if I were myself angry is all the "pretense" needed by simulation theory. Against Gallagher (2007, 361), I cannot see how creating this extra representation (i.e., the mental state attribution) requires the stronger, personal level sense of pretense. Further, this weaker sense of "as if it were I" is sufficient to distinguish it from theory theory accounts, which involve subpersonal inferences from representations of folk psychological generalizations.

Overall, I believe Gallagher loses some important explanatory purchase by confining "simulation" to the personal level. Admittedly, simulation theory developed as a description of conscious simulation, and the personal and subpersonal levels have been inadequately distinguished in many discussions of simulation theory. But denying simulation a subpersonal role seems to obscure the role of states/processes which replicate/resemble those of the perceived target, which I would be in "if it were I" in the position of the target. And it obscures how such

subpersonal processes differ from those involving “theoretical” representations. This contrast seems to be what much of the literature distinguishing theory theory and simulation theory has concerned itself with. The issue is not about the use of the term ‘simulation’. What is important is that the theoretical distinctions made in the literature on simulation theory seem to have some import at the subpersonal level, and should not be lost so as to avoid confusions between personal and subpersonal levels.

Additionally, there remains the question of how personal level, conscious simulation must be characterized at the subpersonal level. It is an open possibility that conscious, imaginative simulation and direct social perception, which are distinct personal level phenomena, are driven by the same or similar subpersonal mechanisms—ones which may be understood in terms of subpersonal simulation. If this were the case, simulation would play a much greater role in our social understanding than Gallagher suggests.

#### **4. Conclusion**

What we are left with is that these phenomenological critics have not provided compelling objections to theory theory and simulation theory as accounts of social perception. We certainly need to respect the phenomenology, by acknowledging a distinction between (a) directly perceiving people’s mental states, and (b) attributing mental states via reflective processes of theorizing and simulation. There is certainly descriptive work left to be done on this front. For example, how are the products of direct perception versus reflective cognition related? Is the direct perception that someone is, say, angry the same as coming to that conclusion via theorizing or simulation? Or is the direct perception of anger different in some way (see note 9).

There is also the issue of distinguishing the various mental state types. Emotions and intentions are most often identified as being directly perceived. But what about states less connected to behavior, like belief? We certainly seem to be able to non-inferentially attribute beliefs to people. How is this different from the direct perception of emotions or intentions?

These are questions befitting phenomenological investigation. But even once we obtain such answers, there remains much about social perception left to account for: namely, the nature of the subpersonal processes enabling this perceptual experience of humans as embodied agents with sensations, emotions, intentions, etc. I have argued that Zahavi and Gallagher's reasons against characterizing these subpersonal processes in terms of theorizing or simulation are unconvincing. Zahavi's arguments are inadequate mainly because he does not directly address what is problematic about the notion of subpersonal theorizing. While Gallagher directly argues against the notion of subpersonal simulation, I have attempted to show why these arguments are unpersuasive.

Perhaps the thrust of Zahavi and Gallagher's criticisms is correct, that we need conceptual development to properly characterize the subpersonal processes underlying social perception. This may very well be the case (see Bechtel 1994, 2005). In this paper I have remained uncommitted about whether theory theory and simulation theory are appropriate subpersonal accounts of social perception, or whether alternative descriptions are needed. But to rule out theory theory and simulation theory as contenders, what is required is a more detailed discussion of appropriate subpersonal explanations than what has been offered thus far by Zahavi and Gallagher.

## Acknowledgements

Earlier versions of this paper were presented at UCSD's Experimental Philosophy Lab and the 2008 Pacific division meeting of the American Philosophical Association. I appreciate the comments provided by those present, particularly Adam Arico. I would also like to thank Bill Bechtel for his comments on this paper and related work.

## Notes

- <sup>1</sup> Since even conscious theorizing must ultimately be characterized subpersonally, the idea must be that the subpersonal processes underlying conscious theorizing are very different from the subpersonal processes enabling direct social perception, such that “theorizing” is never an appropriate characterization of the latter.
- <sup>2</sup> One of the more interesting advances by theory theorists is the characterization of theories in terms of Bayesian networks. This has enabled the creation of computational models of cognitive phenomena such as action understanding (Baker, Tenenbaum and Saxe 2006). While these models remain rather simple at this point, they more precisely characterize the nature of “theorizing” in particular domains, and permit more fine-grained comparisons with human data. Accordingly, they also provide critics such as Zahavi with clearer targets against which to launch objections.
- <sup>3</sup> For more on the role of models in cognition, and the sense in which models are isomorphic to what they represent, see Waskan (2006).
- <sup>4</sup> Grush (2004) explicitly argues for such an account of neural representation, applying it to

motor control, imagery, and perception.

<sup>5</sup> This phrasing is closer to Robert Gordon's version of simulation theory than Alvin Goldman's. Whatever Gallagher's stance on their competing characterizations of simulation, I read Gallagher's objection to subpersonal pretense as cutting against both Gordon and Goldman on subpersonal simulation—see his discussions of Goldman (Gallagher 2007, 361) and of Gordon (Gallagher 2007, 361, n. 10). Below I attempt to undercut Gallagher's objection by appealing to Goldman's (2006) definition of simulation. I do not here delve into the differences between Gordon and Goldman on simulation, or whether Gordon would endorse such a response to Gallagher. I want to thank Marc Slors for his comments on this point.

<sup>6</sup> Gallagher's point here is that the *content* of neural states fails to meet the pretense condition. He also contends that neural mechanisms considered as *vehicles* of representational content fail the pretense condition, writing that “neurons either fire or they do not fire. They do not pretend to fire” (2007, 361). As Gallagher seems to recognize, simulation theory is concerned with matters of representational content more than representational vehicles. Thus I will focus on Gallagher's claim that the content of subpersonal, neural mechanisms cannot meet the pretense condition for simulation theory.

<sup>7</sup> My argument here should not be read as endorsing the standard, simulationist interpretation of motor-based accounts of social perception. There are several different versions of these models, and how they should be interpreted relative to theory theory and simulation theory is not so straightforward. See Herschbach (2008a, 2008b).

- <sup>8</sup> Hybrid theory-simulation accounts, such as Goldman's (2006), do, however, acknowledge a role for representations of folk psychological generalizations. For example, "theory-driven" simulations use theoretical knowledge about the target system to generate appropriate inputs for the simulation process.
- <sup>9</sup> How exactly this resulting state should be characterized is a matter of dispute. Simulation theorists (as well as theory theorists) usually describe this as the production of a *belief* about the other's mental state (e.g., Goldman 2006). But phenomenologists (e.g., Gallagher forthcoming) contrast the "non-conceptual" experience of direct social perception with the "conceptual," belief-based understanding of reflective simulation and theorizing. Although this issue seems important to adequately describing the personal level phenomena of social understanding, it requires delving into the thorny issue of the nature of concepts, and is thus beyond the scope of this paper.

### **Notes on contributor**

Mitchell Herschbach is a Ph.D. candidate in Philosophy and Cognitive Science at the University of California, San Diego. His dissertation integrates research from analytic philosophy of mind, phenomenology, cognitive science, developmental psychology, and neuroscience about human social understanding, particularly our folk psychological capacities and the cognitive mechanisms underlying them.

## References

- Baker, C. L., J. B. Tenenbaum, and R. R. Saxe. 2006. Bayesian models of human action understanding. *Advances in Neural Information Processing Systems* 18: 99–106.
- Baron-Cohen, S. 1995. *Mindblindness: An essay on autism and “theory of mind.”* Cambridge, MA: MIT Press.
- Bechtel, W. 1994. Levels of description and explanation in cognitive science. *Minds and Machines* 4: 1–25.
- . 2005. The challenge of characterizing operations in the mechanisms underlying behavior. *Journal of the Experimental Analysis of Behavior* 84: 313–25.
- Blackburn, S. 1992. Theory, observation, and drama. *Mind & Language* 7, no. 1–2: 187–203.
- Bloom, P., and T. P. German. 2000. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77: B25–B31.
- Churchland, P. M. 1989. *A neurocomputational perspective.* Cambridge, MA: MIT Press.
- Cummins, R. 1996. *Representations, targets and attitudes.* Cambridge, MA: MIT Press.
- Fodor, J. 1983. *The modularity of mind.* Cambridge, MA: MIT Press.
- Frith, U. 2003. *Autism: Explaining the enigma.* Malden, MA: Blackwell.
- Gallagher, S. 2001. The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies* 8, no. 5–7: 83–108.
- . 2005. *How the body shapes the mind.* Oxford: Clarendon Press.
- . 2007. Simulation trouble. *Social Neuroscience* 2, no. 3–4: 353–65.
- . Forthcoming. Direct perception in the intersubjective context. *Consciousness and*

*Cognition*.

- and D. Zahavi. 2008. *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.
- Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Goldman, A. 2006. *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
- Gopnik, A., and A. N. Meltzoff. 1997. *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Grush, R. 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27: 377–442.
- Herschbach, M. 2008a. Control-theoretic models of action understanding: Simulation, theory, or both? Poster presented at the 34th annual meeting of the Society for Philosophy and Psychology, University of Pennsylvania, Philadelphia, PA.
- . 2008b. The concept of simulation in control-theoretic accounts of motor control and action perception. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, ed. V. Sloutsky, B. Love, and K. McRae. Austin, TX: Cognitive Science Society.
- Hurley, S. 2005. The shared circuits hypothesis: A unified functional architecture for control, imitation, and simulation. In *Perspectives on imitation*, ed. S. Hurley and N. Chater, 177–93. Cambridge, MA: MIT Press.
- . 2006. Active perception and perceiving action: The shared circuits model. In *Perceptual experience*, ed. T. S. Gendler and J. Hawthorne, 205–59. New York: Oxford University Press.

- Hutto, D. 2008. *Folk psychological narratives: The sociocultural basis of understanding reasons*. Cambridge, MA: MIT Press.
- Menary, R., ed. 2006. *Radical enactivism: Intentionality, phenomenology and narrative; Focus on the philosophy of Daniel D. Hutto*. Amsterdam: John Benjamins.
- Nichols, S., and S. P. Stich. 2003. *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*.
- Onishi, K. H., and R. Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308, no. 5719: 255–8.
- Perry, J. 1993. *The problem of the essential indexical and other essays*. New York: Oxford University Press.
- Ramsey, W. M. 2007. *Representation reconsidered*. Cambridge: Cambridge University Press.
- Ratcliffe, M. 2006. *Rethinking commonsense psychology*. New York: Palgrave Macmillan.
- Scheler, M. 1954. *The nature of sympathy*. Trans. P. Heath. London: Routledge & Kegan Paul.
- Waskan, J. 2006. *Models and cognition: Prediction and explanation in everyday life and in science*. Cambridge, MA: MIT Press.
- Wittgenstein, L. 1980. *Remarks on the philosophy of psychology*. Vol. 2. Ed. G. H. von Wright and H. Nyman. Trans. C. G. Luckhardt and M. A. E. Aue. Chicago: University of Chicago Press.
- Zahavi, D. 2005. *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, MA: MIT Press.
- . 2007. Expression and empathy. In *Folk psychology re-assessed*, ed. D. Hutto and M. Ratcliffe, 25–40. New York: Springer.