

Mitchell Herschbach

# *False-Belief Understanding and the Phenomenological Critics of Folk Psychology*

**Abstract:** *The dominant account of human social understanding is that we possess a 'folk psychology', that we understand and can interact with other people because we appreciate their mental states. Recently, however, philosophers from the phenomenological tradition have called into question the scope of the folk psychological account and argued for the importance of 'online', non-mentalistic forms of social understanding. In this paper I critically evaluate the arguments of these phenomenological critics, arguing that folk psychology plays a larger role in human social understanding than the critics suggest. First, I use standard false-belief tasks to spell out the commitments of the folk psychological picture. Next, I explicate the critics' account in terms of Michael Wheeler's distinction between online and offline intelligence. I then demonstrate the challenge that false-belief understanding — a paradigm case of mental state understanding — poses to the critics' online, non-mentalistic account. Recent research on false-belief understanding illustrates that mental state understanding comes in both online and offline forms. This challenges the critics' claim that our online social understanding does not require folk psychology.*

## **1. Introduction**

The standard line amongst philosophers and scientists is that humans understand each other by applying their knowledge of 'folk

Correspondence:

Department of Philosophy, University of California, San Diego, 9500 Gilman Drive #0119, La Jolla, CA 92093-0119. Email: mherschb@ucsd.edu

psychology' (FP): that we explain and predict people's behaviour in terms of their mental states. Researchers debate how we go about ascribing mental states (whether through simulation, theorizing, or some combination of the two), and the development of such capacities, but share this characterization of the phenomena constituting human social understanding — namely, that it involves ascribing mental states. This picture has, however, recently come under attack by various authors working in the phenomenological tradition of philosophy, such as Shaun Gallagher, Matthew Ratcliffe, and Dan Zahavi. These *phenomenological critics of FP*, as I'll call them, claim that our social understanding does not primarily involve ascribing mental states. While we can do this in those rare moments when we explicitly reflect on people's behaviour, so the critics argue, the FP-picture fails to apply to the more common episodes of actively engaging and interacting with other people. Their central claim is that most of the time we skillfully navigate the social world without ascribing mental states at all — contrary to the standard line in philosophy and the cognitive sciences.

In this paper I defend the FP-account against this attack from the phenomenological critics. I argue that the critics are right to highlight unreflective, 'online' forms of social understanding, but that they are wrong to restrict FP to reflective, 'offline' cognition. It is true that much of the philosophical and scientific research has portrayed mental state understanding as reflective or 'offline' in nature. The arguments put forward by the phenomenological critics do not, however, prove that all unreflective, 'online' social understanding is non-mentalistic. To expose this limitation in the critics' account, I will focus on what many consider a central aspect of mature FP: false-belief understanding. I will argue that the critics have inadequately addressed this key feature of the FP-picture. By examining the ways in which false-belief understanding can be expressed in our online responses, I show that the phenomenological critics have given too limited a role to mental state understanding in our everyday social understanding.

I will start in section 2 by filling out the FP-account of social understanding, using standard false-belief tasks to illustrate its central features. In section 3, I explicate the phenomenological critics' alternative picture of social understanding, appealing to Michael Wheeler's (2005) distinction between 'online' and 'offline' intelligence. I then show in section 4 the challenge that false-belief understanding poses to the phenomenological critics. I explain why performance on false-belief tasks is best characterized in terms of FP

rather than the non-mentalistic understanding proposed by the critics. Further, I describe recent studies of false-belief understanding in children and adults to illustrate that mental state understanding can be expressed in online as well as offline forms. I use this case study of online false-belief understanding to argue that the phenomenological critics are wrong to restrict mental state understanding to episodes of reflective, offline cognition. While accepting their point that non-mentalistic forms of understanding should be included in accounts of human social understanding, I will argue that FP is much more pervasive in our everyday social understanding than the phenomenological critics claim.

## 2. The Folk Psychological Account of Social Understanding

Most philosophical and empirical discussions of human social understanding over the last few decades have been formulated in terms of ‘folk’ or ‘commonsense’ psychology (e.g., Carruthers & Smith, 1996; Davies & Stone, 1995; Goldman, 2006). Talk of FP is so commonplace now that it may seem a necessary catchall term for whatever knowledge of human psychology and behaviour most of us possess that makes possible human social understanding and interaction. But the standard characterization of FP is not so theoretically uncommitted, instead involving specific assumptions about the nature of interpersonal understanding, and the role of such understanding in social interaction. As Ratcliffe and Hutto (2007) summarize:

the received wisdom about folk psychology encapsulates two chief assumptions: (1) that making sense of actions requires interpreting them in terms of reasons composed of various propositional attitudes (at a bare minimum — beliefs and desires) and (2) that this activity is primarily concerned with providing predictions and explanations of actions. (p. 2)

The two major accounts of FP, *theory theory* (TT) and *simulation theory* (ST), share these basic assumptions, treating FP as involving the attribution of propositional attitudes and other mental states to others in order to explain and predict their behaviour. All forms of TT and most forms of ST treat such mental state ascription as requiring the possession of various mental state concepts, e.g., BELIEF and DESIRE, with mental states conceived of as the inner causes of behaviour. Much ink has been spilled distinguishing TT and ST, with many researchers moving toward hybrid accounts giving roles to both theory and simulation. For most philosophers and scientists, TT, ST, or some combination of the two are the only theoretical options, and the

assumptions they share go unquestioned. FP is presumed to be the proper way to frame any investigation of human social understanding, whether the behavioural experiments of psychologists, neuroimaging studies, or theoretical work of philosophers. It is our proficiency with FP that is offered as an explanation for human social interaction — that without FP, people’s navigation of the social world would be significantly impaired.

These assumptions of FP can be made more concrete by examining a dominant experimental paradigm for studying social understanding: standard false-belief tasks. While I focus on false-belief tasks, as the phenomenological critics have done, many behavioural tasks used to study our understanding of other kinds of mental states (e.g., desire, knowledge, emotion) have the same basic structure as standard false-belief tasks (see, e.g., Wellman & Liu, 2004). One common type of false-belief task involves inducing a false belief in someone by moving an object without their knowledge. In Baron-Cohen, Leslie, and Frith’s (1985) version, often called the *Sally–Anne task*, dolls (named Sally and Anne) are used in place of actual humans as protagonists. The child-participant observes Sally put a marble in a basket then leave the room. While Sally is away, Anne enters and moves the marble from the basket into a box in the same room. Sally then re-enters, and the child is asked, ‘Where will Sally look for the marble?’ The child knows the marble is currently in the box, but must suppress this information, and consider where Sally thinks the marble is located. The child must remember that Sally last saw the marble in the basket and is unaware of Anne’s having moved the marble. Thus the correct response to the experimenter’s question is that Sally believes (falsely) that the marble is in the basket.

Passing standard false-belief tasks such as the Sally–Anne task is taken as signaling a significant development in the acquisition of a FP-understanding of other people. In particular, such understanding requires the child to possess a concept of BELIEF and thereby understand the representational nature of belief, i.e., that people’s beliefs can fail to accord with reality. Having a representational understanding of mental states is crucial to FP’s conception of social understanding, as it allows one to understand people’s actions when they are based on inaccurate representations of reality. Having such understanding can expand the forms of social interaction available to an agent. For instance, it permits the manipulation of other’s representational mental states in order to deceive them, which can be helpful in competitive situations. While proponents of TT and ST differ on what exactly explains the ability to pass false-belief tasks, both take such

tasks to be paradigmatic of our FP social understanding, and success in this domain to be the result of mental state attribution.

Admittedly, false-belief understanding has often been given undue importance in the FP camp — to such an extent that it has sometimes been treated as synonymous with acquiring competence in FP (see Bloom & German, 2000). Of course it must be recognized that other forms of mental state understanding are developmentally more fundamental, and that the development of FP does not end when children can pass standard false-belief tasks around age 4. Further, we do not have much empirical evidence at this point about exactly how important false-belief understanding is to our daily lives, or how easy or difficult it is even for adults. But acknowledging all of this does not detract from false-belief understanding's status as a paradigm case of mental state understanding.

### 3. The Phenomenological Critics of Folk Psychology

Shaun Gallagher, Matthew Ratcliffe, Dan Zahavi, and others working in the phenomenological tradition have offered critiques of ST and TT, and FP in general, and proposed alternative accounts of the nature of human social understanding (Gallagher, 2001; 2005; 2007; Gallagher & Hutto, 2008; Gallagher & Zahavi, 2008; Ratcliffe, 2006; 2007; Zahavi, 2001; 2005; 2007). These phenomenological critics of FP claim that our social understanding does not primarily involve mental state ascription for the purpose of explanation and prediction. The nature of their attack on FP has often been characterized with regard to standard false-belief tasks (Gallagher, 2001; 2005; Ratcliffe, 2007, ch. 4).

FP-accounts take the child's situation in standard false-belief tasks as paradigmatic: they require the child to *observe* someone's behaviour, and *predict* her future behaviour by *attributing* mental states to her, in this case a false belief. If shown Sally unsuccessfully looking for the marble in the basket, the child would be expected to *explain* this behaviour by saying that Sally wanted the marble and falsely believed that it was in the basket. The phenomenological critics point out that standard false-belief tasks place the child in the role of *theorist*, providing conscious explanations and predictions of a third party's behaviour using the concepts BELIEF and DESIRE. Even if the child is not required to provide verbal explanations and predictions, the child must at least somehow report to the experimenter a behavioural prediction, perhaps by pointing to a location. The two dominant FP-accounts, TT and ST, attempt to explain precisely this explicit

explanation and prediction of behaviour based on mental state attribution.

After thus characterizing the nature of the child's cognitive stance in standard false-belief tasks, the phenomenological critics then propose the following criticism: what is not being tested, nor being explained by TT and ST, about the child's behaviour in such tasks is the child's *interaction with the experimenter*. Even if a child is too young to pass standard false-belief tasks, she is still able to talk to the experimenter, understand the experimenter's instructions to observe Sally and Anne, and attempt to answer the experimenter's question about Sally's behaviour. The standard tests of false-belief understanding and other forms of mental state understanding do not investigate the socio-cognitive abilities necessary for such social interaction, where the child is a *participant* in the interaction. The FP-account of social understanding, with its focus on the reflective/theoretical capacities of explaining and predicting behaviour, does not obviously apply to participatory social interactions, even though most FP-accounts make this assumption. While it is possible that theorizing and simulation play a role in such social interactions, tasks placing the subject in the role of *theorist* rather than *participant* will not help much in determining if this is so.

This is a central part of the phenomenological critique of FP as an account of social understanding: the social phenomena for which TT and ST are most clearly explanations are only a subset of the forms of social understanding we humans display; and it is questionable whether FP-accounts developed to explain reflective/theoretical social understanding can be extended to explain participatory social understanding. Behavioural experiments such as the standard false-belief tasks, as well as neuroimaging experiments involving reflective tasks — e.g., reading stories involving people exhibiting various mental states (Gallagher *et al.*, 2000) — that are claimed to provide empirical data regarding our mentalistic understanding of others, are often limited to such reflective phenomena and do not necessarily speak to the understanding required for unreflective social interaction.

The distinction between theoretical and participatory social understanding to which the phenomenological critics want to call attention can be helpfully understood in terms of Wheeler's (2005) distinction between 'online' and 'offline' intelligence (which he introduces while developing a Heideggerian conceptual framework for cognitive science). Online intelligence involves an organism's active sensorimotor engagement with the world: 'A creature displays online intelligence

just when it produces a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli' (p. 12). Offline intelligence, in contrast, is exhibited when an organism is not acting, but reflecting on the world and its possible actions. This is not primarily a contrast between psychological processes that are explicit and available to consciousness, and ones that are not. Rather, it is about the stance an organism takes toward its environment: online sensorimotor interaction versus disengaged contemplation. But the paradigm cases of online and offline understanding clearly do differ in the types of psychological processes involved. Online intelligence, such as skillfully hammering a board, involves little conscious experience: instead of consciously experiencing all of the features of the environment to which you are responding and consciously thinking about how to respond to them, you simply 'cope' with the situation using your implicit know-how. On the other hand, offline reflection more paradigmatically involves conscious experience. To continue the hammering theme, you might consciously imagine possible things to build with the materials you have on hand, and subvocally talk through the pros and cons of each. Of course we sometimes reach such decisions without much conscious reasoning. And surely even paradigmatic episodes of conscious, offline reflection also require many cognitive processes occurring outside of conscious awareness. So the distinction between online and offline intelligence cannot capture all the features relevant to characterizing cognition. But this contrast between online and offline cognitive stances is useful for capturing the distinction in forms of social understanding made by the phenomenological critics.

While standard false-belief tasks require the child to provide a prediction or explanation to the experimenter, and thus involves interaction, FP clearly focuses more on *offline* forms of social understanding, where we are *thinking* about other people's behaviour and making explicit *judgments* about their mental states — in the case of standard false-belief tasks, using mental state concepts to think about someone's false belief to explain or predict their behaviour. What the phenomenological critics want to call our attention to are *online* forms of social understanding, such as the child's active engagement with the experimenter. Gallagher (2005) provides the following list of the sorts of online phenomena the phenomenologists have in mind: 'imitation, intentionality detection, eye-tracking, the perception of meaning and emotion in movement and posture, and the understanding of intentional or goal-directed movements in pragmatic contexts' (p. 230). He calls these 'embodied practices — practices that are

emotional, sensory-motor, perceptual, and non-conceptual' (p. 224). Gallagher's language here is representative of the phenomenological critics' proposal that these capacities for online social interaction are not amendable to description in terms of TT or ST. In other words, the critics suggest that while there are some circumstances in which we make (offline) mental state ascriptions for the purpose of explanation and prediction via theorizing or simulation, the 'embodied practices' constituting online social understanding do not involve mental state attribution, nor explanation and prediction.<sup>1</sup>

These phenomenologists are not alone in recognizing the disconnect between online social understanding and the standard experimental paradigms using offline tasks. For example, Slaughter and Repacholi (2003), in their introduction to a volume on individual differences in 'theory of mind' (i.e., FP), suggest that there are:

several dimensions of difference between laboratory theory of mind tasks and everyday social reasoning ... For instance, in everyday mind reading, we compute mental states *online*, and often *act* on these computations. It would seem rare for us to explicitly *reflect* on the mental state attributions we make in the course of social interactions; instead, we are much more likely to act on those attributions with an *immediate behavioral or linguistic response*. (p. 7, italics added)

While these authors treat 'everyday social reasoning' as mentalistic, they agree with the critics that it is mainly online rather than offline in nature. This view is also making an impact in the neuroimaging literature, where some researchers have begun using online rather than offline tasks of social understanding. These include tasks involving social perception — e.g., perception of people's actions and their expressions of emotion and pain (see, e.g., Decety & Grèzes, 2006) — and approximations of or even actual social interaction (Gallagher & Frith, 2003; McCabe *et al.*, 2001; Oberman, Pineda & Ramachandran, 2007; Schilbach *et al.*, 2006; Spiers & Maguire, 2006). While such research is often cast in the language of FP, the phenomenological critics are clearly not alone in emphasizing online forms of social understanding.

---

[1] In this paper I do not specifically address whether or not 'explanation' and 'prediction' are exclusively offline forms of cognition, as the phenomenological critics claim (e.g., Gallagher, 2005, p. 215).

#### 4. A Challenge for the Phenomenological Critics: False-Belief Understanding

In highlighting the contrast between online and offline forms of social understanding, the phenomenological critics have clearly identified a relevant distinction to which researchers must pay attention, and increasingly are doing so. But their use of standard false-belief tasks to illustrate their objections to the FP-account of social understanding is misleading. Rather than merely marking the online–offline distinction, in their exposition they switch the type of information about other persons being considered. Specifically, they switch from *false-belief understanding* to understanding other properties of persons — e.g., their intentions, emotions, and attentional focus. But the phenomenological critics say little about false beliefs. They clearly want: (a) to deemphasize the significance placed on ‘theory of mind’ experiments such as standard false-belief tasks, since they only focus on offline forms of social understanding; and (b) to place greater importance on online social understanding. But they also claim that online social intelligence should not be interpreted as involving the application of mentalistic concepts for the purpose of mental state attribution. Since the critics’ discussion of standard false-belief tasks does not directly address their views about belief, it is reasonable to ask: what exactly is their view on false-belief understanding? Is the FP-account appropriate in all proposed cases of false-belief understanding, or should performance in such cases be reinterpreted as *not* involving mental state attribution in the way conceived by FP?

The critics certainly argue from considerations of phenomenology that FP-accounts have overstated the importance of conscious mental state attribution. But the cases they use to make this point and present their alternative conception of social understanding do not obviously cover the case of false belief. For example, they propose that we can perceive the emotions and intentions of others in their expressive behaviour — that we do not require theoretical inference or simulative projection from physical behaviour to appreciate such states. Gallagher (2001) illustrates this proposal when writing: ‘The emotional states of others are not, in primary experience, mental attributes that we have to infer. One perceives the emotion in the movement and expression of the other’s body’ (p. 90).<sup>2</sup> But a person with a false belief has an understanding of the world that is *not* shared with the

[2] In other work (Herschbach, 2008), I argue that the phenomenological critics need not treat the ‘direct perception’ of mental states as incompatible with subpersonal level versions of TT and ST.

observer and *cannot* be perceived in their current behaviour. The phenomenological critics also suggest that we often understand people's behaviour in terms of the shared situations and social roles people inhabit, rather than in terms of their mental states. A favourite example (e.g., Ratcliffe, 2007, ch. 4) is that we understand a waiter's actions and can interact with him not because we interpret his mental states, but because we understand his social role as a waiter, in relation to one's own as a customer. Such situational understanding is a matter of understanding how people *normally* act in particular situations with particular roles. But surely our understanding of, say, the waiter role assumes the waiter has true rather than false beliefs (e.g., about what is on the menu, where things are located in the restaurant). The alternative forms of social understanding offered by these phenomenologists thus seem no help in understanding behaviour driven by false beliefs — something humans are clearly able to do. So what should be said on their behalf about false-belief understanding?

Let's start with the online–offline distinction. Surely false-belief understanding comes in not just offline forms, but online forms as well. I can reflect on people's false beliefs, and even communicate these judgments and the behavioural predictions and/or explanations I make from them. But I can also adjust my behaviour in light of what I unreflectively know about people's false beliefs, what could be called *online false-belief understanding*. Do the critics permit characterizing such understanding in terms of FP — i.e., as the online analog of offline false-belief understanding, requiring (amongst other things) the application of a concept of BELIEF? Or is this one of Gallagher's 'embodied practices' which does not involve mental state attribution, and involves a different kind of psychological process? To use some of Gallagher's (2001; 2005) terminology, is what I've been calling online false-belief understanding a form of 'mind-reading' — an understanding of the 'inner,' representational mental states of persons? Or is it a form of 'body-reading' — an 'embodied practice' which does not attribute inner mental states to others, but rather attributes some other form of intentionality more intimately tied to visible behaviour?

#### 4.1. *Cases of online false-belief understanding*

To make the idea of online false-belief understanding more concrete, I'll describe several recent behavioural experiments on false-belief understanding in children and adults. I'll start with two nonverbal tasks used with children as young as 15 months old (Onishi &

Baillargeon, 2005; Southgate, Senju, & Csibra, 2007). Rather than require subjects to verbally or nonverbally make explicit predictions of behaviour based on attributions of false beliefs, these experiments test other ways in which false-belief understanding may manifest itself in children's online behaviour, specifically their looking behaviour. Just as in the Sally–Anne task, these experiments involve false beliefs created by change in the location of an object.

Onishi and Baillargeon (2005) used a violation-of-expectation paradigm to test whether 15-month-old infants have at least a rudimentary understanding of others' false beliefs. After being familiarized with the scene of an agent hiding a toy in one of two locations, then returning later to retrieve the object from that location, infants were shown scenes where the toy was moved without the agent's knowledge. Infants were then presented with the agent searching for the hidden toy either (a) where the agent falsely believed it to be, or (b) where it was actually located. Infants reliably looked longer at instances of (b), the so-called 'unexpected' event, assuming the child expects the agent to search for the toy where she believes it to be located. This experiment tests children's online understanding of others' false beliefs — i.e., children's unreflective expectations about people's behaviour given what children know about their epistemic states — rather than their ability to verbally or nonverbally *report* these expectations to a questioner (or even, seemingly, to themselves).

A problem with looking-time experiments is that they are open to many interpretations about why infants look longer at one condition versus another. As Southgate *et al.* (2007) note, Onishi and Baillargeon's infants might implicitly attribute *ignorance* to the agent rather than a *false belief*. Thus, infants might look longer at the incongruent event (where the agent acts contrary to her false belief) because they do not expect an agent ignorant of an object's actual location to search for it at that location, rather than because they expect an agent to search for an object in the location she falsely believes it to be located. Southgate *et al.* attempted to disambiguate these possibilities using a predictive looking paradigm, where they measured infants' anticipatory eye movements prior to seeing an agent searching for a hidden object. Just as in the previous study, infants were first familiarized with video of an agent watching a toy being hidden in one of two boxes, pausing for a short delay, and then reaching for the toy in that box. In test trials, after the toy was hidden, it was taken out of the box and removed from the scene while the agent was still not looking. This was done to prevent children from being biased in their looking behaviour by knowing the actual location of the toy. The agent then

returned to looking at the two boxes, paused for a short delay, and reached for the toy in one of the two locations: where she believed it to be located (where she saw it hidden), or in the other box where she would have no reason to expect it to be hidden. Using eye-tracking technology, experimenters examined where children first looked after the delay. This served as a measure of where the child expected the agent to search for the toy. Before the agent reached for one of the two boxes, these 25-month-olds more often made their first looks toward, and spent more time looking at, the location in accord with the agent's false belief. Thus, their looking behaviour suggests that the infants expected the agent to look for the toy where she falsely believed it to be located.

Admittedly, these studies do not fully fit the paradigm of 'online' social understanding, since they require children only to passively observe another's behaviour rather than to actually interact with them.<sup>3</sup> Nonetheless, the understanding of false belief required of children seems well characterized as sensorimotor (involving bodily responses to observable stimuli), implicit (not requiring conscious thought) and spontaneous (not requiring explicit instruction from experimenters). It is thus clearly much closer to the 'online' end of the spectrum than the 'offline' end, where we find standard false-belief tasks. Furthermore, it is easy to imagine how the implicit understanding displayed in children's looking behaviour could be extended to cases of actual social interaction. If you can anticipate where a friend with a false belief will look for a desired object, you might help them out by verbally or nonverbally informing them of the object's actual location. While the young children in the above studies may not yet be able to make use of their false-belief understanding in this way, such a response would be of the same general kind as that displayed in these studies, and clearly meet all the criteria for online social understanding.

Other recent studies of false-belief understanding in older children and adults (Carpenter, Call & Tomasello, 2002; Keysar, Barr & Lin, 2003) focus on actual social interactions, and thus serve as examples of full-fledged online false-belief understanding. In these behavioural tasks, participants interpret the speech of an interactive partner in light of their false belief about some feature of the task environment. Importantly, no offline reflection about their partner's mental states is required. Rather, participants must respond online to their interactive partner in a way that requires false-belief understanding.

[3] I want to thank an anonymous referee for pressing me on this point, and for highlighting conversation as an example of online mental state understanding found in many traditional FP-accounts.

Consider first the task given to 3-year-olds by Carpenter *et al.* (2002). Two experimenters (E1 and E2) gave the child two novel objects (A and B) to play with, then taught the child to play a 'hiding' game with them. E1 acted as the 'hider', placing the target object (A) in a container and the nontarget object (B) on the floor to the side of the container, then closing the container. E2 played the 'retriever' role, taking the objects from the container and floor and placing them back in front of the child. In the false-belief condition, E2 left the room, then E1 switched the objects' locations, putting B in the container and hiding A in her bag. E2 returned and tried to retrieve the object in the container, repeatedly using a novel word such as 'toma' to name it — saying things like 'I'm going to get the toma and then we can play with it'. E2 was unsuccessful in opening the container, so E2 and the child instead played with another toy across the room. During this time, E1 placed objects A and B next to each other on a chair. The child was then presented with a retrieval task: E2 noticed the objects, and asked the child to retrieve the object named by the novel word, saying, e.g., 'Oh, look, there's the toma! Can you go get the toma and we'll play with it over here'. To succeed, the child must understand that the novel word names the object E2 falsely believes to be hidden in the container — namely, object A — and use this information to bring E2 the appropriate object. No reflective judgment (e.g., an explicit report of E2's false belief) is asked of the child; instead the child must respond online to E2's request by retrieving the correct toy.

In this study, the child-participant directly interacts with a person holding a false belief, and must understand that person's false belief to successfully negotiate the interaction. Keysar, Lin, and Barr's (2003) study with adults similarly tested the online use of false-belief understanding during verbally-mediated social interaction, using a modified version of Keysar, Barr, Balin and Brauner's (2000) 'referential communication game'. In Keysar *et al.*'s (2000) version, participants sat on one side of a grid containing various objects, with a confederate on the other side playing the role of 'director', instructing the participant where to move objects around the grid. While some of the objects were mutually visible to the participant and the director, others were visible only to the participant. Accordingly, some of the director's instructions were constructed so as to be ambiguous from the participant's perspective, but not from the director's perspective. For example, the director and participant could both see a three-inch-high candle and a two-inch-high candle, but a one-inch-high candle was also visible only to the participant. Thus when the director said to 'Move the small candle to the right', the participant would need to

take into account the director's visual perceptive and knowledge to know that 'the small candle' referred to the two-inch candle (the smallest candle from the director's perspective) rather than the one-inch candle (the smallest from the participant's perspective).

Keysar *et al.* (2003) modified the communication game to test adults' ability to appreciate the director's false beliefs about objects in the grid. They did so by having the participant (out of sight of the director) hide one of the objects — e.g., a roll of tape — in a paper bag and place it in a spot on the grid not visible to the director. The experimenter then misinformed the director about the contents of the bag — e.g., indicating it contained a small ball rather than a roll of tape. The director would then give the participant an instruction that was ambiguous from the participant's perspective but not the director's — e.g., 'Move the tape', when a cassette tape was mutually visible. To determine whether the director's instruction referred to the hidden roll of tape or the cassette tape, participants needed to know that the director had a false belief about the hidden contents of the bag, and thus could not be referring to the roll of tape with the word 'tape'. Participants' understanding of the director's instructions were measured by what object they first looked at and then reached for.<sup>4</sup>

The experimental tasks in Carpenter *et al.* (2003) and Keysar *et al.* (2003) provide evidence of social interactions which require online responses to people's false beliefs. While it is hard to gauge how naturalistic such experimental situations are, they are clear cases where correctly interpreting the speech of an interactive partner requires appreciating their false beliefs about an object relevant to their interaction. Since these tasks involve verbally-mediated social interaction, they undoubtedly fit the paradigm of online social understanding. Participants are not being asked to make reflective judgments about the other person's mental states, or to explicitly predict or explain their behaviour. Rather, false-belief understanding is required to successfully navigate the interaction, to respond online to the other person's verbal request.

In summary, the nonverbal and verbal experimental tasks I have described serve as cases of *online* false-belief understanding because they measure unreflective, spontaneous responses to another person possessing a false belief. While these studies do not address the phenomenological experience of being in these situations, they do not

---

[4] Keysar *et al.* (2003) and other related studies show that adults, and not just children, often fail to take into account other people's beliefs when they diverge from their own. In Keysar *et al.* (2003), participants reached for the hidden object on 22% of false-belief trials.

require participants to engage in offline reflection about others' false beliefs. Rather, false-belief understanding is demonstrated in their online behaviour.

#### 4.2. Possible responses from the phenomenological critics

So how should we interpret such cases? What kind of knowledge about other persons are the participants in these studies displaying? One possibility is the traditional FP-account. Whether they simulate others' mental states or apply theoretical knowledge, the standard FP-picture is that people possess a concept of BELIEF, and are able to make false-belief ascriptions based on their perception of others' behaviour; such false-belief ascriptions are what cause the online responses measured in these studies, as well as the offline, reflective judgments in standard false-belief tasks. For example, one could interpret the infants in the Southgate *et al.* (2007) study as (implicitly) thinking the following: *that person believes the toy is in the box on the left, and he intends to reach for the toy, so I predict he will reach into the box on the left.* These children cannot yet articulate this knowledge verbally, but, on this view, the same knowledge about mental states required for the Sally–Anne task is present in these 1.5- to 2-year-old infants. There is no deficit in children's conceptual knowledge of belief at this young age; that this knowledge manifests itself in their looking behaviour but not other kinds of behavioural responses is attributable to performance deficits.

The verbal studies are a bit more complicated to characterize since language is involved. But the standard FP-account is that language comprehension and production essentially involve mental state attribution: that we interpret people's utterances by inferring their intended meaning, and consider the mental perspective of our audience when speaking (e.g., Sperber & Wilson, 1986/1995; 2002; Tomasello, 2003). In Carpenter *et al.* (2003), for example, the child must recognize that the experimenter falsely believes that the target object is in the container. This is the only way the child could interpret the novel word uttered by the experimenter when attempting to open the container as referring to the target object, rather than to the nontarget object actually located inside the container. According to the FP-account, such understanding again crucially depends on conceptually representing the other person as possessing a false belief.

The phenomenological critics would surely object to such FP-characterizations of online social understanding. But what precisely about them would they reject, and why? Even though they claim that

we can directly perceive emotions and intentions in people's behaviour, and want to treat these states as forms of 'bodily intentionality' rather than representational, 'inner' mental states, the critics surely cannot treat people's false beliefs in the same way. What is interesting about false beliefs is that they are not currently perceivable, and thus paradigmatic of why we treat mental states as 'inner', 'hidden' and distinct from observable behaviour. To understand false beliefs, we must understand that people have points of view on the world which can fail to accord with the world's actual state. It seems that the people in the above studies are indeed responsive to others' *beliefs* — i.e., inner states of epistemic agents serving to represent the world.

One alternative interpretation open to the critics is that people's online behaviour is not actually responsive to other people's false beliefs, but to other properties which are often correlated with their beliefs. Ratcliffe (2007, pp. 53–4), for example, suggests people could solve standard change-of-location false-belief tasks by following a behaviour rule that people look for things where they last saw them. Skeptical challenges like this have repeatedly been put forward against purported behavioural evidence of mental state understanding in nonhuman animals or young children. All researchers must of course respect appeals to parsimony. But these are claims that can be empirically tested — e.g., by designing experimental tasks where simpler, non-mentalistic methods break down. There is not room here to delve into detail about what sorts of behavioural experiments are required to demonstrate genuine false-belief or other forms of mental state understanding (see, e.g., Penn & Povinelli, 2007). But given the widespread acceptance of nonverbal behaviour as a potential indicator of mental state understanding, the phenomenological critics would not be on very firm ground to suggest for this reason alone that *all* purported cases of online false-belief understanding, including the online behaviour of adults, can be accounted for by simpler, non-mentalistic means. The situational contexts and response types, as described in the studies above, are too varied for this general skepticism to hold weight (cf. Call & Tomasello, 2008).

A more substantial objection by Gallagher (2005) and Ratcliffe (2007, pp. 205–1) offers a specific alternative to FP's account of beliefs as inner, representational states: the view that beliefs are 'dispositions to act and to experience in various ways' (Gallagher, 2005, p. 214). Both authors suggest that having a belief does not involve possessing a discrete internal state, but rather that belief attributions can be indeterminate and 'ambiguous even from the perspective of the believer' (Gallagher, 2005, p. 215). It is not clear from these authors'

writings, however, what a dispositional theory of belief is a theory of. Is it an account of what beliefs really are? Is it an account of what everyday people take talk about beliefs to be referring to? Is it an account of what we represent about other people's epistemic states for the purpose of online behaviour? Ratcliffe's (2007, ch. 7) discussion of belief is mostly about the wide range of uses for the term 'belief' in everyday discourse. Gallagher's (2005) discussion wavers, sometimes referring to what we think and talk about other people in understanding their verbal and nonverbal behaviour, and sometimes referring to whether a person 'in reality' has a particular belief.

I find such a view at least relevant to accounts of our *talk* about beliefs — I doubt this discourse is as simple as standard FP-accounts suggest. But I am less satisfied with such an account when attempting to explain the psychological processes by which we track people's epistemic states and act in light of such understanding — i.e., when the focus is on online social understanding. In the tasks described above, we're considering very discrete epistemic states of agents: where they believe a particular object to be located, or what they believe to be found at a particular location. Why not treat these as representational states of agents, and my understanding of these representational states as (meta-)representing them? As discussed above, treating agents as tracking behavioural dispositions is an alternative offered by researchers skeptical of attributing mentalistic understanding. And what exactly is a dispositional account of a *false* belief supposed to look like? How is this distinguished from attributing ignorance to a person? If we restrict ourselves to cases where people are not making explicit, verbal judgments about others' beliefs, as in these behavioural experiments, what kinds of behavioural dispositions could distinguish these cases? Without an explicit analysis of false belief from the phenomenological critics, it is difficult to know what to say here.

The discreteness of the mental states in question also addresses a related objection offered by Ratcliffe (2007, ch. 7): that although researchers in the FP camp describe various situations as involving 'belief understanding', there is not actually a unitary phenomenon deserving this name; people can appreciate a variety of psychological features of other people, so it is unclear what exactly is being tested by experiments of 'belief understanding'. The psychological phenomena Ratcliffe thinks we can distinguish but which are mistakenly lumped together by the FP-account of 'belief understanding' include: sentential attitudes (attitudes directed toward sentences, of the form 'A believes that the sentence 'S' is true') versus propositional attitudes

(attitudes directed toward states of affairs that can be expressed as propositions, of the form ‘A believes that p’); behaviour driven by explicit thought versus habitual behaviour involving no such explicit thoughts; and commitments and convictions that shape our experience, attitudes, and actions in a way distinct from the psychological profile of mere propositional attitudes (e.g., a ‘belief’ in the existence of God). From such cases, Ratcliffe concludes that there is no unitary concept of ‘belief’, and that the FP-account of belief–desire psychology is a misleading characterization of human social understanding.

I agree with Ratcliffe that our understanding of people’s behaviour is not simply a matter of attributing beliefs and desires, i.e., propositional attitudes playing (respectively) informational and motivational roles. This surely oversimplifies the nature of mental state understanding, in both its online and offline forms. But I do not accept Ratcliffe’s conclusion that the FP picture is so oversimplified as to be false, that there are no unitary concepts of belief and desire playing a role in actual human social understanding. More specific to my argument here, whether or not people in the FP camp have overextended the term ‘belief’ does not affect the interpretation of the cases of online false-belief understanding I’ve described. As argued above, these studies concern a well-defined phenomenon: being sensitive to people’s discrete beliefs about particular objects at particular locations. None of the distinctions Ratcliffe makes call into question the unity of this phenomenon, or the FP-account of this phenomenon in terms of appreciating other people’s false representations of the world. How the understanding displayed in these studies relates to other forms of social understanding described in terms of ‘belief understanding’ is an open question. But these experiments are representative of how false-belief understanding is studied experimentally, and show how it can mediate our online social interactions.

Another interpretation open to the phenomenological critics involves the idea that ‘tracking’ false beliefs for the purposes of online behaviour does not depend on conceptually representing them.<sup>5</sup> But when presented in this negative form, this alternative need not stray very far from the FP-account. For example, following up on a discussion of Robert Gordon’s version of ST, Dokic (2002) describes a simulation-based, nonconceptual understanding of belief capable of driving online behaviour. On this view, a concept of BELIEF is required to use the product of a mental simulation to have thoughts or

---

[5] I want to thank another anonymous referee for calling attention to this possible response.

make utterances ascribing a belief to a person. But without the concept BELIEF, a person could still use the results of a simulation routine — i.e., the information about the other's beliefs gained from pretending to have those beliefs (where 'pretending' is not necessarily conscious or explicit) — to drive their behaviour. Such a person shouldn't be said to be making unconscious or implicit belief ascriptions, as the person never entertains thoughts using the concept BELIEF, as required by the standard FP-account. Yet the person is indeed using information about the person's mental states gained from a simulation process. It is possible that online false-belief understanding is driven by such nonconceptual simulation processes, while offline false belief ascriptions involve conceptual representations, as described by the standard FP-account. Another possibility is that nonconceptual simulation characterizes the immature false-belief understanding found in very young children, while adult online false-belief understanding is driven by a conceptual understanding of belief. Either option involves a departure from the standard FP-account that is open to the phenomenological critics. But both retain the core of the FP-account, that we entertain representations of others' mental states.

A more radical account along these lines is that online belief tracking does not involve representing others' beliefs at all. Hutto (2008), for example, develops a nonrepresentational 'biosemiotic' account of the online tracking of others' psychological states. I do not have space here to go into the details of Hutto's view, but one point is especially significant. According to Hutto, language is required to represent the intensional content of propositional attitudes — i.e., that the same object can be represented in different ways or under different descriptions. Accordingly, Hutto believes we can only nonverbally track people's 'intentional attitudes' — i.e., their intentional relations to states of affairs which are nonrepresentational, noncontentful, and extensional (as opposed to intensional) in nature. Intentional attitudes can, for Hutto, be evaluated in terms of their 'success' or 'error,' but not in terms of their truth, as is the case with propositional attitudes. Although he does not explicitly address the purported cases of online false-belief understanding I've described above, Hutto's account would likely characterize them as responding online to people's extensionally 'misaligned' intentional attitudes, rather than to false beliefs understood as propositional attitudes. One reason in favour of this interpretation is that these online false-belief tasks (with the

exception of Keysar *et al.*, 2003<sup>6</sup>) and standard false-belief tasks arguably do not require attending to the intensionality of other people's beliefs (see, e.g., Apperly & Robinson, 2003). They require understanding that another's beliefs can be extensionally off target (e.g., believing that an object is located somewhere it isn't actually located), but not the referential opacity of their beliefs, i.e., that they represent objects under certain descriptions but not others (e.g., that some object is a green ball, but not that it is Sally's favourite toy). Hutto's contrast between intentional and propositional attitudes respects these different aspects of belief understanding. The more controversial part of Hutto's account is his characterization of intentional attitudes as nonrepresentational, noncontentful, and not truth-evaluable, and our understanding of others' intentional attitudes as also exhibiting these properties. We might plausibly deny these claims, and reject Hutto's biosemiotic account of how we are nonrepresentationally sensitive to such states. This would lead us back to the two options we had before: (a) a unified account of beliefs as representational states of persons and a unified account of how we understand them, as in the standard FP-account, or (b) a combination of the conceptual, FP-account with a nonconceptual simulation-based account. But even if we accept Hutto's account of nonrepresentational intentional-attitude tracking as characterizing some purported cases of online false-belief understanding, we need not accept that it covers *all* our online responses. It would certainly be more parsimonious if all online responses involved nonrepresentational tracking of intentional attitudes, leaving an understanding of belief-qua-propositional-attitude to offline reflection. But no convincing argument has been offered that this is case, that propositional-attitude understanding *cannot* drive online responses.

Properly evaluating Hutto's account of nonrepresentational intentional-attitude tracking is beyond the scope of this paper. At this point, it is an avenue which the phenomenological critics could pursue to draw a wedge between online responsiveness to and offline reflection about false beliefs. This interpretation would, however, require the phenomenological critics to concede my point that online behaviour can indeed be driven by *mentalistic* understanding — even if the kinds

---

[6] Understanding the intensional nature of belief is required for some of the conditions in Keysar *et al.*'s (2003) study — e.g., appreciating that a particular object is well described as 'the small candle' from the director's perspective but as 'the middle-sized candle' from the participant's perspective. But not all of their conditions required participants to understand that beliefs can represent objects under particular descriptions.

of mental states at issue are not full-fledged propositional attitudes, as the FP-account contends.

In more recent writings, Gallagher and Zahavi have gone beyond their appeal to non-mentalistic social understanding, directly addressing false-belief and other forms of mental state understanding (Gallagher & Hutto, 2008; Gallagher & Zahavi, 2008). Continuing their critique of standard accounts, they reject FP-style theorizing and simulation as the basis for such understanding. They instead appeal to Hutto's (2008) account of FP understanding as a *narrative* competency which we develop by engaging with others in story-telling practices about people's reasons for action. Given its very recent introduction to the debate, the narrative account has yet to be fully vetted as a genuine alternative to TT and ST. It is important to note, however, that these phenomenological critics continue to treat FP understanding as only necessary in 'puzzling' cases where other non-mentalistic modes of understanding break down (e.g., Gallagher & Zahavi, 2008, p. 193). Yet they have largely left unspecified exactly what counts as a 'puzzling' piece of behaviour. According to the phenomenological critics, is behaviour driven by false beliefs necessarily experienced as 'puzzling'? The online false-belief experiments described above were specifically constructed so as to require an appreciation of another person's false beliefs. But they only measured participants' behaviour, so they do not provide evidence about participants' conscious experience when in such situations. My hypothesis, however, is that at least some of the time we can respond online to people's false beliefs without experiencing their behaviour as puzzling.<sup>7</sup> The burden is on the phenomenological critics to provide an argument that this is not possible. Therefore, even if Hutto's narrative account offers a viable alternative to TT and ST, nothing about it precludes narrative understanding from driving online behaviour in addition to offline reflection.

In summary, I have surveyed several alternatives to the FP-account of false-belief understanding open to the phenomenological critics. None, however, have offered persuasive reason to deny that false-belief understanding is driving the online responses described in section 4.1. These experiments thus provide reason to reject

---

[7] Keysar *et al.*'s (2003) study may speak to this issue. His behavioural results indicate that our initial, automatic response is to attribute to others what we ourselves believe, and that only through a subsequent correction process can we alter this initial egocentric attribution and represent others' beliefs different from our own. That we often actually reach for the wrong object suggests that at least some of the time we consciously notice this error in order to correct for it. Of course this only suggestive, and research directly studying our conscious experience in such situations will be necessary to settle this issue.

Gallagher's (2005) assertion that 'The science of false-belief tests does not provide any evidence for the claim that theory of mind processes are implicit or subpersonal' (p. 219). Admittedly, Hutto's nonrepresentational account of intentional-attitude tracking may be a viable way of treating these online responses as involving less than a full-fledged understanding of belief-qua-propositional-attitude. But even so, intentional-attitude tracking is a form of *online mentalistic* understanding, and thus serves as evidence against the phenomenological critics' claim that online social understanding is purely non-mentalistic. What online responses, in what contexts, are mentalistic or non-mentalistic is an open question — but it is an empirical matter for scientists to address. This case study of online false-belief understanding makes the simpler, conceptual point that the phenomenological critics are wrong to rule out online forms of FP.

## 5. Conclusion

I hope to have exposed a limitation of the phenomenological critics' attack on FP. These experiments of false-belief understanding do support the critics' contention that online intelligence is very important to social understanding, and that many FP-accounts have failed to properly acknowledge this. But the critics have yet to make their case that all online social understanding should be couched in non-mentalistic terms.

How to characterize online versus offline intelligence is a general problem raised by embedded, embodied accounts of the mind, including those based in the phenomenological tradition of philosophy. The FP-account of social understanding is one of the traditional accounts of cognition that must face up to advances made from this perspective. But, if I am correct, FP need not be seen as a relic of traditional cognitive science and philosophy of mind characterizing only a highly restricted range of the phenomena involved in social understanding. The FP-account can be updated to fit within an embedded, embodied approach to social understanding. While mental state understanding may not play the all-encompassing role it has traditionally been assumed to have, it is much more significant than the phenomenological critics allow.

## References

- Apperly, I. A. & Robinson, E. J. (2003), 'When can children handle referential opacity? Evidence for systematic variation in 5- and 6- year old children's

- reasoning about beliefs and belief reports', *Journal of Experimental Child Psychology*, **85**, pp. 297–311.
- Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985), 'Does the autistic child have a 'theory of mind'?', *Cognition*, **21**, pp. 37–46.
- Bloom, P. & German, T. (2000), 'Two reasons to abandon the false belief task as a test of theory of mind', *Cognition*, **77**, B25–B31.
- Call, J. & Tomasello, M. (2008), 'Does the chimpanzee have a theory of mind? 30 years later', *Trends in Cognitive Sciences*, **12** (5), pp. 187–92.
- Carpenter, M., Call, J. & Tomasello, M. (2002), 'A new false belief test for 36-month-olds', *British Journal of Developmental Psychology*, **20**, pp. 393–420.
- Carruthers, P. & Smith, P. K. (1996), *Theories of Theories of Mind* (Cambridge, England: Cambridge University Press).
- Davies, M. & Stone, T. (1995), *Folk Psychology: The Theory of Mind Debate* (Oxford, England: Blackwell).
- Decety, J. & Grèzes, J. (2006), 'The power of simulation: Imagining one's own and other behavior', *Brain Research*, **1079**, pp. 4–14.
- Dokic, J. (2002), 'Reply to Pierre Jacob', in J. Dokic & J. Proust (ed.) *Simulation and Knowledge of Action* (Amsterdam: John Benjamins).
- Gallagher, H. L. & Frith, C. D. (2003), 'Functional imaging of 'theory of mind'', *Trends in Cognitive Sciences*, **7** (2), pp. 77–83.
- Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U. & Frith, C. D. (2000), 'Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks', *Neuropsychologia*, **38** (1), pp. 11–21.
- Gallagher, S. (2001), 'The practice of mind: Theory, simulation or primary interaction?', *Journal of Consciousness Studies*, **8** (5–7), pp. 83–108.
- Gallagher, S. (2005), *How the Body Shapes the Mind* (Oxford, England: Clarendon Press).
- Gallagher, S. (2007), 'Simulation trouble', *Social Neuroscience*, **2** (3–4), pp. 353–65.
- Gallagher, S. & Hutto, D. (2008), 'Primary interaction and narrative practice', in J. Zlatev, T. P. Racine, C. Sinha & E. Itkonen (ed.) *The Shared Mind: Perspectives on Intersubjectivity* (Amsterdam: John Benjamins).
- Gallagher, S. & Zahavi, D. (2008), *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science* (London, England: Routledge).
- Goldman, A. I. (2006), *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (Oxford, England: Oxford University Press).
- Herschbach, M. (2008), 'Folk psychological and phenomenological accounts of social perception', *Philosophical Explorations*, **11** (3), 223–35.
- Hutto, D. & Ratcliffe, M. (ed, 2007), *Folk Psychology Re-Assessed* (New York: Springer).
- Hutto, D. (2008), *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons* (Cambridge, MA: MIT Press).
- Keysar, B., Barr, D. J., Balin, J. A. & Brauner, J. S. (2000), 'Taking perspective in conversation: The role of mutual knowledge in comprehension', *Psychological Science*, **11** (1), pp. 32–8.
- Keysar, B., Lin, S. & Barr, D. J. (2003), 'Limits on theory of mind use in adults', *Cognition*, **89** (1), pp. 25–41.
- McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. (2001), 'A functional imaging study of cooperation in two-person reciprocal exchange', *Proceedings of the National Academy of Sciences*, **98**, pp. 11832–5.

- Oberman, L. M., Pineda, J. A. & Ramachandran, V. S. (2007), 'The human mirror neuron system: A link between action observation and social skills', *Social Cognitive & Affective Neuroscience*, **2** (1), pp. 62–6.
- Onishi, K. H. & Baillargeon, R. (2005), 'Do 15-month-old infants understand false beliefs?', *Science*, **308** (5719), pp. 255–58.
- Penn, D. C. & Povinelli, D. J. (2007), 'On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind"', *Philosophical Transactions of the Royal Society Series B*, **362**, pp. 731–44.
- Ratcliffe, M. (2006), "'Folk psychology' is not folk psychology", *Phenomenology and the Cognitive Sciences* **5**, pp. 31–52.
- Ratcliffe, M. (2007), *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind and Simulation* (Basingstoke, England: Palgrave Macmillan).
- Ratcliffe, M. & Hutto, D. (2007), 'Introduction', in Hutto, D. & Ratcliff, M. (2007).
- Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., et al. (2006), 'Being with virtual others: Neural correlates of social interaction', *Neuropsychologia*, **44** (5), pp. 718–30.
- Slaughter, V. & Repacholi, B. (2003), 'Individual differences in theory of mind: What are we investigating?', in B. Repacholi & V. Slaughter (ed.) *Individual Differences in Theory of Mind: Implications for Typical and Atypical Development* (New York: Psychology Press).
- Southgate, V., Senju, A. & Csibra, G. (2007), 'Action anticipation through attribution of false belief by two-year-olds', *Psychological Science*, **18** (7), pp. 587–92.
- Sperber, D. & Wilson, D. (1995), *Relevance: Communication and cognition* (Oxford, England: Blackwell, 1986).
- Sperber, D. & Wilson, D. (2002), 'Pragmatics, modularity and mind-reading', *Mind & Language*, **17** (1–2), pp. 3–23.
- Spiers, H. J. & Maguire, E. A. (2006), 'Spontaneous mentalizing during an interactive real world task: An fMRI study', *Neuropsychologia*, **44** (10), pp. 1674–82.
- Tomasello, M. (2003), *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Cambridge, MA: Harvard University Press).
- Wellman, H. M. & Liu, D. (2004), 'Scaling of theory-of-mind tasks', *Child Development*, **75** (2), pp. 523–41.
- Wheeler, M. (2005), *Reconstructing the Cognitive World: The Next Step* (Cambridge, MA: MIT Press).
- Zahavi, D. (2001), 'Beyond empathy: Phenomenological approaches to intersubjectivity', *Journal of Consciousness Studies*, **8** (5–7), pp. 151–67.
- Zahavi, D. (2005), *Subjectivity and Selfhood: Investigating the First-Person Perspective* (Cambridge, MA: MIT Press).
- Zahavi, D. (2007), 'Expression and empathy', in D. Hutto & M. Ratcliffe (ed.) *Folk Psychology Re-Assessed* (New York: Springer).

Paper received January 2008