Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/101366

Please be advised that this information was generated on 2020-09-21 and may be subject to change.

The Model-Model of the Theory-Theory

Why 'theory of mind' seems ubiquitous, even though it isn't

Marc Slors

Abstract

"Theory of mind" (ToM) is widely held to be ubiquitous in our navigation of the social world. Recently this standard view has been contested by phenomenologists and enactivists. Proponents of the ubiquity of ToM, however, accept and effectively neutralize the intuitions behind their arguments by arguing that ToM is mostly subpersonal. This paper proposes a similar move on behalf of the phenomenologists and enactivists: it offers an explanation of the intuition that ToM is ubiquitous that is compatible with the rejection of this ubiquity. According to this explanation, we use ToM-talk mostly to model or reconstruct nonmentalizing social-cognitive processes for practical and theoretical purposes. The intuition that ToM is ubiquitous is the result of mistaking the model for the real thing.

"Efficient practice precedes the theory of it."
[Gilbert Ryle, *The Concept of Mind*, London: Hutchinton House, 1949, p. 30]

"Theory of Mind", or "ToM" for short, is "the cognitive achievement that enables us to report our propositional attitudes, to attribute such attitudes to others, and to use such postulated or observed mental states in the prediction and explanation of behavior."[1] (Garfield et. al. 2001: 494) In philosophy, psychology, linguistics and neuroscience, the majority view has it that ToM is ubiquitous in our navigation of the social world. Since about a decade, however, this standard view is no longer universally accepted. A growing number of philosophers, mainly of a phenomenological or enactivist bent, contend that our primary and most pervasive way of making sense of each other in daily social interaction is not ToM-driven (Gallagher 2001, 2004; Hobson 2002; Hutto 2004, 2008a, 2008b; Ratcliffe 2007; Hutto and Ratcliffe 2007; Gallagher and Zahavi 2008a; Zahavi 2005, 2007;

[1] Throughout this paper I will remain neutral on the question what such a cognitive achievement looks like in detail. The usual understanding of ToM is in terms of rules or laws, but according to Maibom's (2003) account ToM consists of psychological models we employ while assuming background hypotheses. Nothing in this papers hinges on this.

Bermúdez 2003; Maibom 2007). Many (though not all[2]) of these philosophers claim that ToM is used only in the relatively scarce cases where more basic social cognitive skills that *are* claimed to be pervasive do not suffice. ToM-adherents now start to launch counterattacks on the phenomenologists and enactivists (Herschbach 2008a, 2008b, Currie 2008, Spaulding 2010), i.e. there is an emerging debate over the ubiquity of ToM. The issue is whether the social cognitive mechanisms we use most frequently in social interaction should be understood in terms of the application of a ToM. Thus, the debate is between extremes. Either ToM is claimed to be ubiquitous, or it is downplayed as a fringe phenomenon.

This paper starts from the observation that, given the structure of the debate, a stable position in it requires not only a defense of one of these extremes, it also requires that the intuitions in favor of the other extreme be explained or explained away. There is no use in trying to convince a flat-earth believer that the world is round, unless one explains at the same time why it seems flat. The "ToM-ist" orthodoxy has a well entrenched solution to this problem: it can be conceded that from a phenomenological point of view it *seems* as if we rarely use our ToM in social navigation, when the operations of our ToM's are thought to take place mainly at an unconscious, sub-personal level. Within the phenomenologist and enactivist camps, by contrast, no explanation has been offered for the ToM-ist intuition that we understand most or all intentional actions of others in terms of motivating beliefs and desires.

The goal of this paper is to propose an explanation of the intuition that ToM is ubiquitous that fits the enactivist/phenomenologist position. This explanation hinges on the idea that ToM is not only used as a social cognitive *mechanism*, but also as a *model* for non-mentalizing social cognitive mechanisms. These non-mentalizing mechanisms are in fact ubiquitous, as enactivists and phenomenologists claim. The intuition that ToM is ubiquitous, according to this explanation, is the result of mistaking the model in terms of which we understand the ubiquitous way we understand others for the real thing.

The paper is set up as follows. In Section 1 I will outline the reasons for contesting the ubiquity of ToM and indicate how ToM-ists accommodate these reasons by conceiving of ToM mainly as a sub-personal affair. I will discuss serious problems with this move and briefly sketch the enactivist proposal to do without sub-personal ToM. I will argue that even though these considerations do not count as a decisive rejection of the ubiquity of

[2] Ratcliffe, for instance, appears to reject ToM entirely as a philosophers myth. Hutto, on the other hand, rejects only the theoretical nature of the capacities required for daily interaction. But his substitute for ToM, the practice of providing "folk-psychological narratives", is not at all claimed to be a rare social phenomenon.

(sub-personal) ToM, they certainly warrant an investigation into the feasibility of a phenomenological/enactivist position that rejects the ubiquity of ToM. A remaining serious obstacle for such a position is the lack of an explanation for the apparently widespread intuition that ToM is ubiquitous. In Section 2 I will sketch the "model-model" of ToM as a means of accommodating the intuition that ToM is ubiquitous without contradicting the enactivist/phenomenologist rejection of ToM. In Section 3, I will defend the model-model of ToM by showing that, contrary to first appearances, it offers the best interpretation of theorizing about ToM acquisition in developmental psychology.

1. The Contested Ubiquity of Theory of Mind

According to a majority view in social cognition research, "(…) our basic grip on the social world depends on our being able to see our fellows as motivated by beliefs and desires we sometimes share and sometimes do not." (Currie and Sterelny 2000: 145) This view is so dominant and intuitive—at least to those steeped in the mainstream literature of social cognitive neuroscience, linguistics and analytical philosophy—that it may look as if there is no alternative. In the words of Baron-Cohen, "(…) it is hard for us to make sense of behavior in any other way than via the mentalistic (or 'intentional') framework (…). [A]ttribution of mental states is to humans as echolocation is to the bat. It is our natural way of understanding the social environment." (Baron-Cohen 1995: 3-4)

But it is precisely this "naturalness" of ToM that is currently being questioned. One important complaint is that a ToM-based conception of social cognition assumes an unnatural observer model of social interaction. ToM-based accounts of social interaction hinge on the idea that understanding and predicting the behavior of others is key (an idea that was attacked earlier; see Morton 1996). And the picture of what such understanding and prediction amounts to is very much that of a detached observer adopting a third-person perspective on the interpreted other. Either we are thought to infer the beliefs and desires of the other person from observed behavior by applying our theory of mind. Or we are thought to ascribe the outcomes of our own decision making procedures to the observed other after we put ourselves in their mental shoes. In real life, however, most of the time we do not observe and predict each other, we *interact* with each other adopting an engaged, second-person perspective. In Hutto's words, "Understanding others is not essentially a spectator sport" (Hutto 2008b: 12).

With this emphasis on the dominance of the second-person, I-you, perspective in social interaction, it becomes much less natural to consider the attribution of beliefs, desires, intentions, hopes and fears etc. to play the central role they are usually assigned. Compare the way in which we understand our own actions (cf. Gallagher 2001: 88). I understand my own actions at a pragmatic intentional level as being goal-directed. This does not usually involve my consciously ascribing beliefs and desires to myself. When I am thirsty and reach for a glass of water, I am aware of the purpose of that action. But such awareness does not usually involve explicitly ascribing the desire for water to myself, or the belief that water quenches thirst. The same is true, according to phenomenologists, of understanding the actions of others in second-person interaction: we understand their actions in their contexts in terms of purposefulness and goal-directedness, without ascribing more abstract and generalizable propositional attitudes.

Instead of attributing beliefs and desires, Gallagher, Hutto and many others claim, we are directly responsive to intentionality, purpose and goal-directedness in the contextualized actions of others, just as we directly see emotions in facial expressions and gestures (see also Hobson 2002, Gallagher and Zahavi 2008, Hutto and Ratcliffe 2007, Morton 2003). When someone walks up to me at a coktail party, purposefully, with an outstreched hand, I immediately 'see' her intention to shake hands. I need not make any inference about the mindset of this person for that (see also, Hutto 2008a: 6). Similarly, we directly perceive emotions in the facial expressions of others. To use one of Gallagher's favoratie quotes:

> (...) [W]e 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 outstreched 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 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 certainly is no sensation of another person's mind (...) I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts. (Scheler 1954: 260; see also Wittgenstein 1980: § 570)

How do we know that the blushing we see in someone's face is indeed an expression of shame and not of, say, excitement or physical exercise? Here phenomenologists and enactivists typically invoke the crucial role of contextual setting and co-occurrence of various actions, gestures and facial expressions. Blushing combined with avoidance of the gaze of a person who has just made a critical remark about the blusher, for instance, is not easily mistaken as a sign of excitement or physical exercise.

There are three important elements in the phenomenological/enactivist rejection of the ubiquity of ToM. 1. First of all, the claim is that for most day to day interpresonal interactions we need not and do not ascribe propositional attitudes—full-blown beliefs and desires with a specifiable propositional content—to others. It is sufficient to ascribe purposefulness, goal directedness and intentionally on a pragmatic level. In Hutto's terminology, it is sufficient to ascribe *intentional attitudes* (to be distinguished from *propositional* attitudes; see Hutto 2008a ch. 3). To the extent that such intentional attitudes are mental, this means that we do understand others as minded beings. 2. However, this does not mean that it is granted that we ubiquitously interpret the behaviour of others in terms of a hidden mental realm. The claim is that in most day to day interactions, the basic intentional states of others are not hidden behind the actions, gestures and expressions for which they are merely causally responsible. Rather, they are present *in* these actions, gestures and expressions, and we perceive them as such. 3. This means that ascription of intentional attitudes proceeds noninferentiality. It is not the case that we postulate unobservable states behind the behaviour of others. Nor is it the case that we infer a specific interpretation of someone's actions, gestures or expressions by using contextual cues. Rather, the claim is that we directly perceive meaning in contextualized (combinations of) actions, gestures and expressions.

For these reasons and in this sense, our daily navigation of the social world is claimed not to require our ubiquitously wielding a *theory* of mind. Intentional attitudes are not the (folk-psychological) kinds of mental state that figure in ToM. Moreover, since they are noninferentially observable, no theory is required to postulate them. This obviously constitutes a rejection of the so-called 'theory theory' (TT) of social cognition as an account of most day to day personal interactions. But it also implies a rejection of a dominant version of the so-called simulation theory (ST) according to which we understand others not by wielding a theory but by putting ourselves in the other's mental shoes—as it is often put. For most versions of ST agree that navigation of the social world involves (1) ubiquitously ascribing propositional attitudes to others, where such attitudes are taken to be (2) nonobservable states that are at best causally responsible for observable behaviour of others (Goldman 2006, Nichols et. al. 1996). (3) In many cases ST can also be said to involve inference when it comes to the final ascription of propositional attitudes after the simulation procedure or when it comes to initiation such procedures through the generation of 'pretend beliefs and desires' (see also Perner 1996). Having said this, though, I should also stress that an important version of ST does not meet these criteria (Gordon 1986, 1996). This version is often explicitly not targeted by the

phenomenologist/enactivist critics (Hutto 2008a: 138-139) and takes itself to be congenial to the phenomenologist/enactivist position (Gordon 2008).

Despite their rejection of the ubiquity of ToM, phenomenologists and enactivists do concede that ToM abilities are, every now and then, necessary when social perception fails. But such occasions are exceptional, according to Gallagher (2008: 165): "(…) mentalizing or mindreading are, at best, specialized abilities that are relatively rarely employed, and they depend on more embodied and situated ways of perceiving and understanding others, which are more primary and pervasive." Hutto (2008a) argues that instead of mentalizing in a ToM-ist fashion we employ "folk-psychological narratives" to supplement what he calls "unprincipled bodily engagements." Ratcliffe is more radical in thinking that folk-psychology, even if it is understood in non ToM-ist fashion such as proposed by Hutto, is a philosophers construct: "(…) what is labeled an "everyday", "commonsense" or "folk" psychology and routinely accepted as the core of human social life is actually nothing of the sort and bears little relation to how people understand each other." (Ratcliffe 2007: 2) At any rate, according to these critics ToM is not, as Baron-Cohen would have it, "our natural way of understanding the social environment." At best it is an exceptional phenomenon in social interaction.

The response of proponents of the ubiquity of ToM to these considerations is simple and appears, at least initially, effective. Proponents of the ubiquity of ToM do not dispute phenomenological claims about the apparent absence of ToM in most social interactions. Instead, they argue that these claims leave completely open the option that ToM is ubiquitous at the sub-personal level. Thus, referring to ToM in terms of its two main guises, the theory theory and the simulation theory, Herschbach writes:

> I agree that [Gallagher and Zahavi's] phenomenological claims have bite at the personal level, distinguishing direct perception from conscious theorizing and simulation. Their appeals to phenomenology and other arguments do not, however, rule out theory theory and simulation theory as accounts of the subpersonal processes underlying social perception. (Herschbach 2008b: 223)

In a similar vain, Spaulding writes:

> With mindreading, there is a process (theorizing or simulating), and there is a product (an explanation or a prediction). In general, neither the process nor the product need be consciously accessible, let alone phenomenologically transparent. If only the product of mindreading (the explanation or prediction of behavior) is available at the conscious level, then presumably this would feel phenomenologically as if our interactions are the result of immediate, non-mentalistic understanding. (Spaulding 2010: 131)

This move is all but ad hoc. The idea that ToM is mainly a sub-personal affair is a majority view with a respectable tradition independent of the present discussion. Especially in the theory-theory camp few philosophers claim that we mainly theorize consciously about the mental states of others. In the simulation theory camp there are versions that reject conscious simulation (e.g. Gordon 1995, 1996) while the ones that do allow for conscious simulation leave ample room for sub-personal simulation (cf. Goldman 2006, ch. 6).

Even though the move is not ad hoc, it does not go down well with the enemy. Gallagher and Zahavi, notably, object to the notion of sub-personal ToM. Their misgivings initially coincided partly with what Blackburn (1992) called "the promiscuity objection": characterizing sub-personal processes as theoretical would stretch the meaning of the term "theory" beyond its normal use (cf. Zahavi 2005: 181). According to Gallagher (2005: 215), the term 'theory' and it's associated notion of 'explanation' are normally taken to imply reflective consciousness. If so, it is not clear what is being said when sub-personal, unconscious, processes are considered to be theoretical. But when pressured, their hostility to sub-personal theorizing turns out to be of a kind with their rejection of personal level theorizing. The fact is that attempts have been made to meet the promiscuity objection. Herschbach (2008b: 227-8) points this out, citing Gopnik and Melzoff (1997) as an example, who define "theory" in terms of structural, functional and dynamic features, leaving consciousness or reflection out of their definition. In a joint paper, Gallagher and Zahavi answer that they do and did recognize the existence of definitions of "theory" that avoid promiscuity. They mention explanatory and predictive power, counterfactual projection, the introduction of unobservable entities, and the integration of information within a small number of general principles as proposed defining features. They then point out that it is in particular the postulation of unobservable entities that makes theorizing unfit to characterize social cognition (Gallagher and Zahavi 2008: 238), both at the personal and the sub-personal level.

Gallagher and Zahavi's objection to the idea of sub-personal ToM, then, only bears on one aspect of it. The point is that even if one theory-defining element—the postulation of unobservable entities—is left out, it still seems perfectly intelligible to speak of sub-personal processes as instantiating a theory. Features such as counterfactual projection and integration of information within a relatively small number of principles might well suffice to think of ToM in sub-personal terms without rendering the term "theory" vacuous. All in all, then, Herschbach and Spaulding's idea to "go sub-personal" in order to defend the ubiquity of ToM against phenomenologist considerations appears intelligible and sensible. This move allows them to

explain why they think ToM can be ubiquitous despite appearances to the contrary. The strength of "going sub-personal" is precisely in the fact that the phenomenological view of social interaction is respected—and neutralized.

But this does not mean that the going subpersonal move suffices to save the ubiquity of ToM. For one principal argument against the ubiquity of ToM stems not from phenomenological considerations, but from considerations about computational parsimony that apply just as much to the sub-personal level. As Bermúdez notes:

> (...) the application of [ToM] principles requires identifying, among a range of possible principles that might apply, the ones that are the most salient in a given situation. It requires identifying whether the appropriate background conditions hold, or whether there are countervailing factors in play. It requires thinking through the implications of the principles one does choose to apply in order to extrapolate their explanatory/predictive consequences. (...) [This] certainly makes them rather unwieldy. (Bermúdez 2003: 31-2)

Many social interactions are simply not complex enough to warrant the use of such an unwieldy cognitive mechanism. From the viewpoint of computational parsimony, then, this observation counts as a serious disadvantage of the view that ToM is ubiquitous.

Spaulding interprets the charge in the context of the contrast between the phenomenology of social interaction and the sub-personal processes that are at play in this and responds by noting (Spaulding 2010: 135-6): "Of course what happens at the sub-personal level is going to be computationally more complex than how it seems to us at the personal level, but that is no strike against theories about sub-personal processes. (...) The contrast of the computational complexity of mindreading with the phenomenal ease and instantaneousness of social interaction is not, in and of itself, evidence that mindreading cannot be our normal way of understanding others." One thing to note about this reply is that, as a counter-argument, this puts an impossible and unreasonable burden of proof put on the anti-ToM-ist. In the absence of reasons to rule out in principle a ToM-ist reading of sub-personal processes, Spaulding will not be convinced. But that is to treat ToM-ism as the default position without argument. More importantly, neither Bermúdez nor anyone else will deny that the sub-personal processes underlying simple social interaction are complex. The issue is not just about complexity, it is about *unnecessary* complexity. Bermúdez' point is that conceiving of neural processes underlying simple interactions in ToM terms makes them much more complex than they need to be.

Here it is crucial to consider the competition. According to the anti-ToM camp, daily social interaction is not facilitated by the attribution of propositional attitudes, but rather by the perception of basic intentional attitudes, goal directedness, emotions etc. The neural mechanisms at play in such social perception are for a large part uncharted territory. But this does not mean that nothing can be said about them. Gallagher and Zahavi (2008a: 178-179) think social perception can be modelled on enactive theories of perception as sensory-motor processes (cf. e.g. Noë 2004; see Gallagher 2009 for an elaborate defense of this view). Typically, mirror neuron activity and phenomena of 'empathic resonance' (di Pellegrino et. al. 1992, Gallese 2001) are viewed as important contributors to such processes of enactive social perception (see also Hutto, forthcoming). This means that a simulationist interpretation of them is (largely, but not entirely—see Slors 2010) resisted. The point here is that such neural mechanisms, although no doubt very complex, avoid the *extra* complexity that Bermúdez observes to be involved in ToM mechanisms. Thus, Spaulding's argument against Bermúdez fails, and considerations about computational parsimony do seem to count against the idea that ToM is ubiquitous.

A close kin of the problem of computational complexity is the frame problem. It is sometimes claimed (e.g. Heal 1996) that conceiving of our access to the minds of others in theoretical terms runs into the following difficulty: there is no *a priori* limit on the information that possibly bears on the ascription of mental states to others from a purely theoretical, non-empathic point of view. Spaulding claims that enactivists and phenomenologists run into the same problem since they "must explain how it is that we can determine which facial gestures, eye movements, expressive movements, etc. are relevant to understanding other people (…). There is no *a priori* limit on what embodied cues are relevant to understanding others." (Spaulding 2010: 136-7)

This charge misses its goal, however. The point is that from an enactivist point of view we don't need *a priori* limits on cues relevant to understanding others, as long as the sensori-motor processes underlying social interaction are wired so as to respond to the relevant ones. If they do, they "embody knowledge" about the relevance of behavioral cues. But since they do not *represent* that knowledge *as such*, talk of *a priori* limits on cues is beside the point. Wondering how sensori-motor systems manage to "select" the relevant information, "discarding irrelevant information", is like wondering why it is that the moon has exactly the right mass and speed not to either crash into the Earth or fly away into space. Much is presently unknown about the details of these sensori-motor processes, that is certainly

true. But there is no frame problem (see Dreyfus 2006 for further arguments to this effect).

Hence, considerations of computational complexity and the frame problem count strongly against the idea that ToM is ubiquitous, even if ToM is mainly or exclusively conceived at the sub-personal level. Although such considerations are admittedly not conclusive, they do warrant a further investigation into the feasibility of the phenomenologist/enactivist view of day to day social interaction without ToM. One major obstacle for this view—the one I shall be concerned with in the remainder of this paper—is the fact that so far enactivists and phenomenologists have done nothing to explain and/or explain away the immensely widespread intuition that ToM *is* ubiquitous in daily social interaction.[3] Where ToM-ists *are* able to take into account the fact that phenomenologically speaking ToM seems not ubiquitous (by going sub-personal), anti-ToM-ists have no parallel argument to allow for the fact that an overwhelming majority of philosophers and psychologists consider ToM ubiquitous. My aim in the following two sections is to provide such an argument.

2. ToM as a Model versus ToM as a Mechanism

The aim of this section is to explain why ToM is widely taken to be ubiquitous, assuming that enactivists and phenomenologists are right in claiming this is not, in fact, the case. A good way to introduce the idea is to draw a partial parallel with Daniel Dennett's intentional stance theory (Dennett 1987). According to Dennett the reality of beliefs and desires is exhausted by the fact that our behavioral patterns can easily and usefully be tracked through adopting the intentional stance. Adopting the intentional stance towards a system is understanding that system's behavior as being issued by beliefs and desires. Some have described this as an "as-if" theory of mental state ascription (McCulloch 1990). But on the intentional stance theory beliefs and desires are more than just fictions. They are *useful* fictions that track *real* behavioral patterns that cannot in any other way be tracked (Dennett 1991a). Still, beliefs and desires do not exist as psycho-neural realities, according to the intentional stance theory. There really are no semantically evaluable internal causes of actions that accord with our belief-desire psychology (cf. Fodor 1985: 78) in the observed agents (see however, Slors 2007). On Dennett's account that would be a "gratuitous bit of misplaced concreteness" (Dennett 1987: 55).

[3] Hutto (2009a, 2009b) is in some respects an exception. A comparison of Hutto"s explanation of the ToM intuition and the one provided in the next section, however, is beyond the scope of this paper.

The intentional stance theory is presented as a more or less ToM-ist position.[4] According to it we navigate the social world mainly through attributing beliefs and desires. As such the position is of no use to the project of explaining the intuition that ToM is ubiquitous against the background assumption that ToM is not ubiquitous. My proposal, however, is to take the intentional stance idea one level up: the cognitive mechanisms involved in understanding others and navigating the social world can best be tracked and understood for practical purposes in terms of the attribution of beliefs and desires, i.e. the application of a ToM. But that is not to say that these cognitive mechanisms use or implement a ToM. ToM is a *model* of ubiquitous social cognitive mechanisms that would otherwise be intractable. Since the model provides our ways of thinking and talking about such mechanisms, it may easily be mistaken for the real thing. This explains the intuition that ToM is ubiquitous. Let me go over this proposal in a little more detail.

We sometimes explain or excuse our actions towards other by stating what it is we, perhaps mistakenly, thought the other was thinking or what she or he wanted or hoped or feared, etc. In the context of a common history and certain shared norms of conduct, the following kind of explanation or excuse is common: "I'm sorry, I thought you wanted *x* because you said *y* and so I figured you would probably wanted me to.....". In this type of explanation we model our own grasp of someone else's conduct as theorizing in terms of attributed beliefs, desires and other propositional attitudes. Applying Dennett's move to his own theory: we talk about ourselves *as if* we adopted the intentional stance. But the reality of this stance, the reality of our application of a ToM, is often similar to how Dennett envisages the reality of beliefs and desires. Beliefs and desires, in his theory, are not psycho-neural realities in the heads of the people whose behavior we predict and explain in their terms. Likewise, according to the proposed position our adopting the intentional stance, our wielding a ToM is—very often (see below for the exceptions)—not a psycho-neural reality. In that sense, it is very often a fiction. But it is a *useful* fiction that models *real* social-cognitive mechanisms that are otherwise intractable. Thus it is incorrect to say that ToM is entirely a philosopher's myth (cf. Ratcliffe 2007).

The claim that our wielding a ToM is often not a psycho-neural reality might need some further elaboration. In this context it is useful to

[4] The intentional stance theory is ToM-ist in the sense that Dennett writes as if he considers the ascription of beliefs and desires ubiquitous. But although he sometimes seems to defend a theoretical reading of what it is to adopt the intentional stance (e.g. 1987, chapter 4), at other occasions he rejects such a reading (e.g. Dennett 1991b).

distinguish between different levels at which a cognitive system—in this case a social cognitive system—can be described. One can describe the social cognitive mechanisms at play in our daily interaction at the neural level, at the functional level and at the phenomenological level. The claim defended here is that wielding a ToM does not describe what is going on in the social cognitive processes that underlie most daily interpersonal interaction at the phenomenological level, nor does it describe these processes at the neural level. This is in line with the phenomenologist/enactivist position discussed in the previous section. At the functional level of description, however, wielding a ToM may be said to capture these processes. But only if a functional description is not taken to involve identifying a network of interrelated states definable in terms of their causal roles. For that would imply that the occupiers of these roles, which are neural states, may be described as implementing a ToM mechanism. If, however, a functional description is taken to refer to a description that captures, to some degree of accuracy, what a system *does* in repsonse to the inputs it receives, that is, if a functional description is an interpretive description of a system *taken as a whole*, it seems reasonable to say that our daily social cognitive mechanisms can be functionally described as wielding a ToM. This boils down to saying that we can *model* the neural mechanisms and the phenomenal experiences at play in daily social interaction in terms of our wielding a ToM.

Thus, I claim that when we give explanations or excuses such as "I thought you wanted me to .....", this is not usually intended as a phenomenologically accurate recounting of mental episodes. There is usually not a moment at which one literally thinks to oneself "x wants me to do y." Likewise, there is not, usually, a conscious inference preceding such a thought. But even when the pick-up of intentions, emotions and wishes from facial expressions, gestures, voice intonations and, obviously, the contents of another person's speech proceeds in a few split seconds, we are often able to give a relatively detailed reconstruction of this pick-up in ToM-like terms when we feel we need to provide one. Here's an example: You arrive at a party, famished. The first thing you see is a table stacked with food. The way you look at the table probably betrays the state of your stomach. The host looks at you, he smiles, raises his eyebrows and gives a brief nod towards the table while retaining eye contact. He means to signal "looks good, eh." So when you reach for a sandwich, he says, sharply: "wait, not everyone's here yet." Then you apologize: "I'm sorry. I thought saw that you could see I was hungry and that you meant I could have one." You did not literally think that. Your response to the nod was in all likelyhood too quick to allow for a conscious inference from perceiving your host's nod to the ascription of a belief about the state of your stomach which, in

conjunction with his having pity on you, may have caused his nod. Nevertheless, that thought does capture your unreflective assessment of the situation. Here it could typically be argued that such reconstructions are sub-personal ToM routines made explicit. But 'sub-personal' can either mean 'neural' or 'functional' here. On the phenomenologist/enactivist position explored here, your pick-up of your host's intention may well be a matter of being responsive to his intentional attitudes, in which case the 'neural' interpretation of 'sub-personal' would be wrong-headed (see previous section). If, on the other hand, 'sub-personal' is interpreted as 'functional', making sub-personal ToM routines explicit boils down to affirming the model-model of ToM. In that case, the reconstruction is what Dennett calls a "heuristic overlay". That is, it may well be that the ToM explanation does not mirror an in principle specifiable, tractable sub-personal process, but instead employs a psychological vocabulary that serves the purpose of *inter*personal sense-making rather than *intra*personal description.

A couple of things are worth noticing at this point. First, this interpersonal sense-making may take the form of providing explanations or excuses such as the above mentioned. But it may also serve other functions, such as explaining to a third party why one thinks *x* did such and such or decided so and so, for instance in a dispute over someone's motivations. Secondly, it is important to note that the fact that we *can* model social cognitive processes in ToM terms does not imply that we *do* this very often. Finally, the social cognitive processes that are reconstructed in ToM terms may be diverse. In the phenomenological literature (and in the remainder of this paper), there is an emphasis on the direct pick-up of intentions and emotions. But understanding others based on character traits, moods, social roles (Ratcliffe 2007, Goldie 2007, Morton 2003) or narratives (Hutto 2008a) is also often described in non-mentalizing terms. Many instances of these ways to understand others can be modeled in ToM-like ways too.

Thus, the proposal is that we often use ToM to reconstruct or interpret our non ToM-driven social cognitive mechanisms being either implemented at the neural level or present at the phenomenological level. It may be instructive to compare this proposal with the idea that we understand our own mental states not via introspection but by applying our ToMs to ourselves via what is known as 'first-person mindreading' (Carruthers 2009). According to the proposal under discussion, we interpret our own responses to others in ToM terms in circumstances in which we need to explain our direct assessment of others. This may sound like a form of first-person mindreading too. Indeed, I would not object to the term, nor would I object to expanding the idea to include all conscious self-ascriptions (see below). There is, however, a crucial difference between Carruther's

notion of first-person mindreading and the current proposal: Carruthers thinks of first-person mindreading mainly in sub-personal, unconscious terms. Thus, he is a proponent of the idea that ToM is ubiquitous, not as a model of sub-personal social-cognitive mechanisms, but as a sub-personal cognitive mechanism itself.

The comparison of the current proposal with Carruthers' notion of first-person mindreading raises two questions: 1. To what extent is Carruthers' ToM-ist notion of first-person mindreading susceptible to the general criticism levelled against ubiquitous ToM in the previous section? 2. Carruthers' notion of first-person mindreading hinges on the idea that we are naturally endowed with a sub-personal mindreading module, but what can say about the nature and origins of the kind of first-person mindreading involved in the model-model of ToM? Let me briefly sketch the contours of the answers to both questions.

1. I take it that first-person mindreading along ToM-ist lines is not immune to the general criticism of ubiquitous ToM outlined in the previous section. The considerations about computational complexity and the frame problem as reasons to favor a more parsimonious enactivist view on the neural mechanisms underlying the bulk of our daily social interactions may apply as well to issues of self-attribution. For here too, applying ToM means determining which principles are relevant, which background conditions hold, which countervailing factors are at play, etc. Whether the phenomenologist/enactivist view involves a computationally more parsimonious picture, depends on what it offers as an alternative for Carruthers' ideas about the sub-personal mechanisms underlying self-attribution. Gallagher (2000, 2004a, 2005) offers a 'neurophenomenological' account of the sub-personal processes involved in self-ascription of intentions and thoughts, taking his cue from earlier comparator models (cf. Wolpert et. al. 1995; Frith et. al. 2000). There is no space to go into the details of his proposal here. But the crucial aspect of his theory for the present discussion is that it avoids the extra computational burden of invoking theory, just like a sensory-motor theory of social perception avoids theory in the case of third-person attribution. Therefore, Bermúdez' criticism applies to day to day first-person mindreading as well.

2. In order to say something about the nature and origins of our use of ToM as a model, it is important to note the congeniality of the model-model and theories that view ToM primarily as a socio-linguistic phenomenon, summarized as follows by Astington (see also Nelson et. al. 2003; Nelson 2005; Dunn and Brophy 2005; Hutto 2008a):

> Bruner (1983) proposes that parents treat infants' spontaneous gestures as intentional communications and thus infants come to see themselves as having

> intentions and start to communicate intentionally. In a similar way, parents talk to toddlers about their thoughts, feelings and desires, and the children come to see themselves as holding such states. Parents also use the same linguistic terms to talk about other people. That is to say, the children's own experience is construed in the same terms that are applied to others, and they come to see that others have similar experiences to their own. Thus, linguistic development is fundamental to the acquisition of mental state concepts, because without language the child would not learn about these concepts, which are in the speech practices of culture. In this sense the theory of mind, perhaps even mind itself, is a cultural invention. (Astington 1996: 187-8)

On this view, children learn learn to construe their own mental lives and that of others in terms of the psychological framework of the culture in which they are socialised. [5] This presupposes that children have an initial non-ToM-based grasp of the desires, intentions and emotions of themselves and others. For in order to be able to learn how to apply a psychological vocabulary, one needs to have a grasp of that to which this vocabulary is applied. Thus, Astington implicitly presupposes that children have some grasp of the mental lives of themselves and others *prior to* their ability to apply the ToM terminology. There is an extensive developmental psychological literature on how children acquire this initial grasp of minds. There is no room to discuss this literature here, but it should be stressed that this literature is at least compatible with (and more probably: suggestive of) the phenomenologist/enactivist outlook on our basic social cognitive abilities (see Gallagher 2004b; Gallagher 2005 ch. 9 and Hutto 2008a, ch. 3.).

Hence, the view that ToM development is socially and linguistically scaffolded and the view that the initial basic grasp of others is non-mentalizing complement one another. Their combination explains the sense in which ToM can be seen as not just as a "heuristic overlay", but also as an "expository overlay". To use a metaphor, ToM is a socio-cultural mold in which the wax of our non-mentalizing grasp of others acquires the shape and structure that allow us to put it to use in linguistically mediated social contexts.[6] The idea of the model-model is that these 'linguistic transformations' (Hutto 2008a, ch. 4) involve not just the nonconceptually

[5] It is not clear to what extent Astington applies the generality constraint here (Evans 1982), implying that mental predicates are understood only if they can be applied to both self and others.

[6] What is left out of this account is the notion that our socio-culturally shared ToM vocabulary also shapes and regulates the development of the mental states children ascribe to themselves and others as well as the kinds of social behaviour that give rise to the ascription of propositional attitudes. Such 'mind-shaping' (Zawidsky 2008) would provide an explanation, not so much of the ubiquity of ToM, but of ToM-interpretable social behaviour. The notion of mind-shaping is compatible with the model-model of ToM.

grasped mental states of others and ourselves, but *also this initial nonconceptual grapsing itself*. Reconstructing our initial grasp of the actions of others makes our understanding of others (and ourselves) tractable. And that serves social purposes such as explaining or excusing ones actions towards others (or epistemic purposes such as gaining explicit self-knowledge).

So, one type of use of ToM is to provide models of non-mentalizing social cognitive mechanisms. But once we have acquired the ability to forge such models, the possibility arises that it be used autonomously, in hypothetical mode, to predict and understand others. ToM can be used in abstraction from and/or as substitute for our non-mentalizing social cognitive abilities. The need for such use of ToM may arise when more basic social cognitive mechanisms are not sufficient to make a certain situation transparent to us. Suppose you see a car stop in front of a bank. Two armed and masked men rush out of the car and into the bank. The driver remains in the car, leaving the motor running. We don't need to mentalize in order to see that there is a bank robbery going on. No implicit or explicit ascription of beliefs and desires is necessary. Some background knowledge and contextualizing of actions is sufficient. But now suppose we see the same scene in front of a tourist office. Now a non-mentalizing understanding is insufficient. The situation must be made transparent by a mentalizing hypothesis: these men probably think the tourist office is a bank.

Obviously there are more complex situations in which we are required to use our ToM, not to model our social cognitive mechanisms, but as a social cognitive mechanism in its own right.[7] Every now and then we need to reason about someone's motives. The point made by phenomenologists, however, is that such occasions are relatively scarce. In day to day life our non-mentalizing social cognitive mechanisms usually do the bulk of the work. To repeat Gallagher: "(…) mentalizing or mindreading are, at best, specialized abilities that are relatively rarely employed (…)." It is worth noting that he goes on to say that mentalizing and mindreading—i.e. the use of ToM as an autonomous social cognitive tool—"depend on more embodied and situated ways of perceiving and understanding others, which are more primary and pervasive." (Gallagher 2008: 165). This dependence, about which Gallagher says little, can be explained by recognizing that ToM development is a socio-linguistically scaffolded activity that necessarily makes use of pre-existing non-mentalizing social cognitive abilities.

[7] I use the word "mechanism" very loosely here. The idea is that ToM may serve as an epistemic tool. Obviously, if used as such, on the view I expound in this section, ToM is a personal-level conscious affair.

Thus, to summarize the current proposal, there are two ways in which ToM is used. It can be used to model basic, non-mentalizing social cognitive mechanisms, which serves certain social purposes. Or it can be used as a social cognitive mechanism in its own right. Neither of these uses is ubiquitous. But the first use ensures that we naturally talk and think about our ubiquitous non-mentalizing grasp of others in terms of ToM. The intuition that ToM is ubiquitous, then, arises because the model in terms of which we talk and think about social cognition is quite understandably mistaken for the real thing.

3. The Model-Model of ToM in Developmental Psychology

On the model-model of ToM, the use of ToM that is responsible for the intuition that ToM is ubiquitous, is making sense of our ubiquitous but more basic sense-making of others. This meta-sense-making typically is a personal level, conscious process. At first glance, then, the model-model appears to be contradicted by the dominant view in developmental psychology. There is a near consensus in this area about the idea that children as young as two or three years old are *mentalizing* or at least *theorizing*—very probably unconsciously or at a sub-personal level—when they pass what is known as nonverbal or implicit false belief tests. At first glance, this near-consensus appears to contradict the model-model of ToM.

In this section, however, I will argue that the contradiction is merely apparent on a plausible reading of the nature of theorizing in developmental psychology. I will argue that the developmental psychologist's depiction of early social cognitive skills of young children in terms of theorizing, mentalizing or both can best be viewed as the scientific variant of our commonsense modeling described in the previous section. On this interpretation the debate on infant social cognition in developmental psychology confirms rather than contradicts the model-model.

In order to appreciate the current debate on nonverbal false belief tests we need to start with the traditional, verbal version of the false belief test, instigated by Dennett's comments (amongst others) on a groundbreaking paper on the ToMs of chimps (Premack and Woodruff 1978, Dennett 1978). The test, initiated by Wimmer and Perner (1983; see also Baron-Cohen et. al. 1985) and replicated in endless varieties, hinges on the transfer of an item (a ball, a piece of chocolate, etc.) from container A to container B, unseen by the agent who originally put it in A. Children who watch the transfer scene are asked where the agent will look when she returns to get her item. Consistently, research shows that (at least in most cultures) on average children will produce the correct answer—i.e. the agent

will look in the box she originally put it in, since she doesn't know it has been transferred and hence falsely believes it is where she put it—from the age of 4 onwards (Wellman et. al 2001).

Recently, however, nonverbal tests involving similar transfer scenarios have been devised which suggest, according to researchers, that children much younger than 4 years have a rudimentary grasp of others' false beliefs. Originally, Onishi and Baillargeon (2005) used a violation of expectation paradigm, measuring looking times of children. After the transfer scene, children were presented with two further scenarios, one in which the agent looks inside the box where she put the object, and one in which the agent looks inside the box where the object actually is. Children as young as 15 months tend to look longer at the latter scenario, which is interpreted as a sign of surprise. In other words, apparently they expected the agent to look where she falsely believes it to be. In order to avoid the objection that looking times are multi-interpretable, Southgate et. al. (2007) used eye-tracking technology and a predictive looking paradigm (I will skip the details for the sake of brevity since their experimental set-up is more sophisticated). They found that "25-month-old infants correctly anticipate an actor's actions when these actions can be predicted only by attributing a false belief to the actor." (Southgate et. al. 2007: 587). More nonverbal tests have been devised with similar results, some even claiming to find false belief understanding in children as young as 13 month olds (Surian et. al. 2007).

These results appear to be paradoxical. Children are deemed to grasp the fact that other people have false beliefs while they will be unable to express this fact for some time. Worse, their verbal utterances will at times contradict their looking behavior, which appears to betray a grasp of false beliefs. Consider, for instance, the following example by Perner (who, by the way, thinks children acquire an initial sense of false belief no earlier than by the age of 3). Referring to an earlier study (Clement and Perner 1994) he writes:

> Wendy Clements tested children on the traditional, unexpected transfer test of false belief understanding: Sam the Mouse puts his cheese in one of two boxes which are placed in front of the two mouse holes at the extreme top corners of the display. While Sam is inside his sleeping quarters invisible to the child, the cheese is transferred to the other box. When Sam wakes up and wants his cheese, children are asked where he will look for it. Children's responses show the traditional developmental picture (...). Almost all three year olds (2 years 11 months to 3 years 2 months) answer with the second box (actual location of the cheese) while most four year olds get it right, answering with the empty box (where Sam believes the cheese is). What is new is that Clements recorded where children *looked* when Sam's desire for cheese was mentioned. A surprising 80% of the three year olds looked at the empty box. Importantly, only a few of the very young

> children looked there in a control condition in which the only difference was that Sam knew where the cheese was because he saw the cheese being moved before he went to bed. (Perner 1996: 98).

At this point it is important to note that from a thoroughgoing enactivist point of view, there is no paradox. What the child learns in the course of its psychological development, according to enactivism, are primarily ways of interacting with the world, including other people. From that perspective, what the child is able to *do* when it passes a nonverbal false belief test is radically different from what it can *do* when it passes a verbal false belief test. In order to pass a nonverbal false belief tests the child has to display spontaneous responses driven by the kinds of sensori-motor process that tracks the intentional attitudes of others (Herschbach 2008b: 44). Such responses are likely to subserve social interactions in real life such as complex forms of what Gallagher (2004b) and Hobson (2002) call secondary intersubjectivity, involving joint attention and joint intentionality (think e.g. of how this looking behavior may elicit responses from others through gaze-tracking). What children must be able to do in order to pass the verbal false belief test, by contrast, is something quite different. They have to be able to participate in a small conversation about an observed scenario in which they allow a second person to gain accurate knowledge of the future behavior of one of the protagonists in that scenario. Put in these terms, the abilities required for passing both tests are so different that no paradox arises from the fact that at some stage children can pass one test but not the other.

The paradox does arise, however, when we turn to cognitive explanations behind these abilities. Being able to enlighten others about the future behavior of a third party in a small conversation and being able to exhibit certain looking responses are both claimed to be explained in terms of knowledge of another person's false belief. Without further qualification of what such knowledge consists in, then, a puzzle emerges. If passing both tests requires the same kind of knowledge, then why is it that three year-olds pass one but not the other?

The solution here is to differentiate between degrees or levels of false belief understanding or even between different systems of false belief understanding. The very young child's understanding of false belief may be similar but nevertheless distinct from the 4 year old's understanding. Perner and Ruffman (2005), for instance, think that Onishi and Baillargeon's 15 month-olds do not really need a ToM to pass the test:

> [W]e acknowledge their suggestion that infants expect the observed person to act in a particular way. However, we propose that this can be based on behavior rules. Infants may have noticed (or are innately predisposed to assume) that people look for an object where they last saw it and not necessarily where the object actually is.

> Again, such a rule captures something implicit about the mind, because the rule only applies as a result of the mind mediating between seeing and acting. Nonetheless, infants can simply know the rule without any conception that the mind is the mediator. (Perner and Ruffman 2005: 215)

The idea, then, is that instead of a ToM, the young child uses a "theory of behavior" (ToB) to pass the nonverbal false belief test. Such a ToB may consist in behavioral rules, but it may also consist of person-object-place associations leading to simple predictions of behavior. The same sorts of claims have been made about the social cognitive skills of chimps (Povinelli 2001, Povinelli and Vonk 2004).

Trading ToMs for ToBs strikes many as being too behaviorist, however (Csibra and Southgate 2006; see e.g. Tomasello et. al. 2005 for a parallel view regarding chimps). But this leaves us with the paradox. Apperly and Butterfill (2009), however, explicitly aim to solve the paradox by arguing for the idea that we are endowed with two separate ToM systems. Drawing a parallel with number cognition, they contend that we have one system for computationally efficient but inflexible mindreading and another system for flexible but cognitively demanding mindreading. The idea, then, is that passing a nonverbal false belief test requires a *minimal* theory of mind (MToM; see also Butterfill and Apperly, forthcoming), while it is likely that engaging in conversation about the future behavior of others requires the flexible full-blown version of ToM. Thus, introducing a MToM besides a regular ToM solves the paradox: 3 year olds have the former but not the latter. In a similar fashion other theorists speak of a "naïve theory of mind" (e.g. Bogdan 2009) or "early mindreading skills" (Nichols and Stich 2003).

It is impossible, without going beyond the scope of this paper, to go into the details of the various experiments and the ToM, ToB or MToM that are being proposed as explanations for their outcomes. But enough has been said to explain the idea that the debate between ToM, ToB, MToM and other interpretations of the cognitive mechanisms at play in passing nonverbal false belief tests is a debate over the best reconstruction or model of these mechanisms. Four features of debate are important:

First, in order to draw conclusions about ToM, ToB, or MToM capacities from nonverbal false belief tasks the behavioral responses of infants are not simply taken at face value. In experiments based on the violation of expectation paradigm, the primary explanandum is expectation. In experiments based on the predictive looking paradigm, the primary explanandum is prediction. Thus, the infants' looking responses are only explained indirectly. The point here is not that interpretations of the infants' looking responses in terms of violation of expectation or prediction are far fetched. They aren't. The point is that the looking responses are

interpreted in terms that go beyond the direct, immediate, unreflective responses they are to the infants themselves. Remember that Astington observes, in the quote in the previous section, that "parents treat infants' spontaneous gestures as intentional communications," which signals the start of a process in which children are scaffolded into our mentalistic language game. Here something similar is the case. Scientists give a "thicker" reading of the infants' responses than children themselves are able to give.

Secondly, the ideas of a ToB or a MToM are somewhat ad hoc. Again, this is not intended as a negative qualification. It is merely intended to point to the fact that the idea of such "proto-ToMs" did not develop as a theoretical consequence of the ToM-ist conception of social cognition as such. It developed in response to experimental results that were unexpected, at least from the reigning point of view that ToM abilities are acquired by the age of 4. As a result, both ToB and MToM are tailor-made to fit the results of nonverbal false belief tests (see especially Perner and Ruffman 2005, and Butterfill and Apperly, forthcoming).

Thirdly, so far the choice between ToB and MToM (or some other option) is underdetermined by the data from nonverbal false belief tests. Of course such tests are being improved precisely to limit the number of possible interpretations of their outcomes. Thus, for instance, Southgate et. al.'s (2007) experimental set-up is an improvement of Onishi and Baillargeon's original experiment, designed specifically to rule out interpretations of the infant's response in terms of ignorance or priming by the verbal instructions of the experimenter. Still, the set-up does not rule out in principle an interpretation of the infant's looking response in terms of sensitivity to person-object-place associations acquired in familiarization trials. Neither does it rule out an interpretation in terms of behaviour-based rules.

The question is whether this underdetermination of the theory of infant social cognition by the data is a sign of underdeveloped experimental design or whether it is principled. There are good reasons to assume the latter. The crucial difference between ToM/MToM and ToB is that the former does and the latter does not invoke mental mediation between seeing and acting (in the character in the false belief scenario). As Perner and Ruffman indicate in the above quote, the idea of ToB instead of ToM does not question mental mediation as such. The idea is that the relations between seeing and acting relevant for behavior prediction are tractable for the infant without invoking mental mediation. The issue between ToB and ToM/MToM is when mental mediation is likely to be invoked in order to keep the relations between seeing and acting tractable enough to yield correct behavioral predictions. Thus, it is implicitly agreed that both ToB and ToM explain the data. The question is which is more likely or

"realistic". It may be the case that further experiments may help to tip the balance of likelihood or explanatory usefulness in one direction or the other (see Butterfill and Apperly, forthcoming for suggestions). But the evidence will never be such as to rule out one option *in principle*.

Fourthly, it is important to be clear about the explanatory "level" at which the debate takes place. Although data about neural activity play a significant role in the debate (e.g. in Perner and Ruffman 2005 and Apperly and Butterfill 2009), these are not data that point to the implementation of a specific cognitive architecture without—usually well argued-for—*interpretation* in terms of ToB, ToM or MToM. The bulk of the debate is really about inferences to the best explanation of behavioral data. Here "explanation" refers to cognitive design, to hypothesized principles of information processing. And "best" is a pragmatic and instrumental notion the measure of which appears to be a balance between computational parsimony and fit with a description of children as natural "mentalizers".

Each of these four features of the debate on nonverbal false belief tests suggests in its own way that theorizing about infant social cognition in terms of ToM, MToM or ToB is basically reconstructive modeling. When these features are taken together, this suggestion is nigh impossible to resist. The three "theories" do not really function, in the debate, as detailed hypotheses about the actual neural circuitry underlying social cognition. Rather, they function as abstract descriptions of cognitive design. In other words, the debate is not over the truth but over the explanatory and predictive usefulness of the three options. This is entirely in line with the model-model of ToM. The issue is not whether the infant's expectation of future behavior as manifested in looking behavior is being *caused by* either belief- or behavioral rule understanding, rather the issue is whether it is best *analyzed* in such terms.

Thus, the idea is that ToB, MToM and ToM can best be viewed as models that derive their claim to reality not from correspondence with isomorphic brain processes or the conscious experiences of children in test situations, but from explanatory and predictive usefulness. To be sure: the same would hold for a possible enactivist interpretation of the results of implicit false belief tests.[8] Such an interpretation would have to prove itself as having equal explanatory potential while being more parsimonious than ToM, ToB or MToM interpretations (see section 1). Obviously there is no room to argue this point on behalf of enactivism here. The purpose of this

[8] No worked-out enactivist account of these experiments has been proposed as far as I know. I believe that Butterfill and Apperly's (forthcoming) account of MToM in many respects comes close—with a few modifications—to being an enactive account (see de Bruin, Strijbos and Slors, forthcoming).

section is merely to emphasize the model-status of ToM or ToM-like interpretations of implicit false belief tasks, so that the phenomenologist/enactivist claim that ToM is not ubiquitous at the phenomenological and/or neural level is shown to be compatible—despite initial appearances to the contrary—with recent developments in mindreading research.

Viewing ToM, ToB, and MToM as models is not in any way intended to undermine the status of developmental psychology. It is merely to make a claim about the nature of explanations in developmental psychology. There is nothing scientifically suspect about model explanations (think e.g. of models used to predict the weather or economical or demographic models). Better models with more predictive or explanatory leverage signal an increase in knowledge. The point of arguing for the idea that ToM functions as a model in developmental psychology is to emphasize that the enactivist/phenomenological rejection of the ubiquity of ToM is not contradicted by current developmental psychological research and theorizing. And again, like in the previous section, it allows us to explain the intuition that ToM is ubiquitous without granting that it is. The alleged ubiquity of ToM is derived from the ubiquity of the social cognitive processes that can be modeled in terms of ToM. Again, the ubiquity assumption is the product of mistaking the model for the real thing.

Acknowledgements

Thanks to Derek Strijbos, Dan Hutto, Fleur Jongepier, Leon de Bruin, Birgit Knudsen, Bas Donders, the audience of the Engaging Minds workshop at the University of Hertfordshire and two anonymous referees for this journal for helpful comments.

References

Apperly, I.A. and Butterfill, A.B. 2009: Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116 (4), 953-970.

Astington, J. 1996: What is theoretical about the child's theory of mind?: a Vygotskian view of its development. In Carruthers, P. and Smith, P.K. (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 184-199.

Baron-Cohen, S. 1995: *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge MA: MIT Press.

Baron-Cohen, S., Leslie, A. and Frith, U. 1985: Does the autistic child have a 'theory of mind'? *Cognition*, 21, 37-46.

Bermúdez, J.L. 2003: The domain of folk psychology. In O'Hear, A. (ed), *Minds and Persons*. Cambridge: Cambridge University Press, 25-48.

Blackburn, S. 1992: Theory, observation and drama. *Mind and Language*, 7: 187-203.

Bogdan, R. 2009: *Predicative Minds: The Social Ontogeny of Propositional Thinking*. Cambridge MA: MIT Press.

Butterfill, A.B. and Apperly, I.A. forthcoming: Minimal theory of mind.

Carruthers, P. 2009: How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences* 32: 121-182.

Clement, W.A, and Perner, J. 1994: Implicit understanding of belief. *Cognitive Development*, 9, 377-395.

Csibra, G. and Southgate, V. 2006: Evidence for infants' understanding of false beliefs should not be dismissed. *Trends in Cognitive Science*, 10 (1), 4-5.

Currie, G. 2008: Some ways to understand people. *Philosophical Explorations*, 11 (3), 211-218.

Currie, G. and Sterelny, K. 2000: How to think about the modularity of mind-reading. *Philosophical Quarterly*, 50, 145-160.

De Bruin, L.C., Strijbos, D.W. and Slors. M.V.P. forthcoming: Mindreading in early social development?

Dennett, D. C. 1978: Intentional systems in cognitive ethology: The Panglossian paradigm defended. *The Brain and Behavioral Sciences*, 6, 343-390.

Dennett, D.C. 1987: *The Intentional Stance*. Cambridge MA: MIT Press.

Dennett, D.C. 1991a: Real patterns. *The Journal of Philosophy*, 88 (1), 27-51.

Dennett, D.C. 1991b: Two contrasts: Folk-craft versus folk-science, belief versus opinion. In J. Greenwoord (ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science*. Cambridge: Cambridge University Press, 135–148.

di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese and G. Rizzolatti 1992: Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91: 176-80.

Dreyfus, H.L. 2006: Overcoming the myth of the mental, *Topoi*, 25, 43-49.

Dunn, J. and Brophy, M. 2005: Communication, relationships, and individual differences in children's understanding of mind. In Astington, J. and Baird, J.A., *Why Language Matters for Theory of Mind*. New York: Oxford University Press, 50-69.

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

Fodor, J. 1985: *The Modularity of Mind: An Essay on Faculty Psychology*. Ca,bridge MA: MIT Press.

Frith, C. D., Blakemore, S., & Wolpert, D. M. 2000: Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Brain Research Review*, 31(2-3): 357–363.

Gallagher, S. 2001: The practice of mind: Theory, simulation of primary interaction? *Journal of Consciousness Studies*, 8 (5-7), 83-108.

Gallagher, S. 2000: Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4: 14–21.

Gallagher, S. 2004a: Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology*, 37(1): 8–19.

Gallagher, S. 2004b: Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. *Philosophy, Psychology and Psychiatry*, 11, 199-217.

Gallagher, S. 2005: *How the Body Shapes the Mind*. New York: Oxford University Press.

Gallagher, S. 2008: Inference or interaction: social cognition without precursors. *Philosophical Explorations*, 11 (3), 163-174.

Gallagher, S. 2009: Neural simulation and social cognition. *Contemporary Neuroscience* 6: 1-17.

Gallagher, S. and Zahavi, D. 2008a: *The Phenomenological Mind*. London: Routledge.

Gallagher, S. and Zahavi, D. 2008b: The (in)visibility of others: a reply to Herschbach. *Philosophical Explorations*, 11 (3), 237-244.

Gallese, V. 2001: The 'shared manifold' hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies* 8 (5-7): 33-50.

Garfield, J.L., Peterson, C.C. and Perry, T. 2001: Social cognition, language acquisition and the development of the Theory of Mind. *Mind and* Language, 16, 494-541.

Goldie, P. 2007: There are Reasons and Reasons. In Hutto, D.D. and Ratcliffe, M. (eds), *Folk Psychology Reassessed*. Dordrecht: Springer, 103-114.

Goldman, A. 2006: *Simulating Minds: the philosophy, psychology and neuroscience of mindreading*. New York: Oxford University Press.

Gopnik, A. and Melzoff, A.N. 1997: *Words, Thoughts, and Theories*. Cambridge MA: MIT Press.

Gordon, R.M. 1986: Folk-psychology as simulation. *Mind and Language*, 1: 159-71.

Gordon, R.M. 1995: Simulation without introspection and inference from me to you. In.

Gordon, R.M. 1996: 'Radical' simulationism. In Carruthers, P. and Smith, P.K. (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 11-21.

Gordon, R.M. 2008: Beyond Mindreading. *Philosophical Explorations*, 11 (3): 219-222.

Heal, J. 1996: Simulation, theory and content. In Carruthers, P. and Smith, P.K. (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 75-89.

Herschbach, M. 2008a: False-belief understanding and the phenomenological critics of folk psychology. *Journal of Consciousness Studies*, 15 (12), 33-56.

Herschbach, 2008b: Folk psychological and phenomenological accounts of social perception. *Philosophical Explorations*, 11 (3), 223-236.

Hobson, P. 2002: *The Cradle of Thought*. London: Macmillan.

Hutto, D.D. 2004: The limits of spectatorial folk-psychology. *Mind and Language* 19, 548-573.

Hutto, D.D. 2008a: *Folk Psychological Narratives: The Socio-Cultural Basis of Understanding Reasons*. Cambridge MA: MIT Press.

Hutto, D.D. 2008b: The Narrative Practice Hypothesis: clarifications and implications. *Philosophical Explorations*, 11 (3), 175-192.

Hutto, D.D. 2009a: Lessons from Wittgenstein: Elucidating folk psychology. *New Ideas in Psychology*, 38.

Hutto, D.D. 2009b: ToM Rules, but it is not OK. in A Costall (ed.) *Against Theory of Mind*. Basingstoke: Palgrave.

Hutto, D.D. and Ratcliffe, M. 2007: *FolkPsychology Reassessed*. Dordrecht: Springer.

Maibom, H. 2003: The mindreader and the scientist. *Mind and Language*, 18 (3), 296-315.

Maibom, H. 2007: Social systems. *Philosophical Psychology*, 20 (5), 557, 578.

McCulloch, G. 1990: Dennett's little grains of salt. *The Philosophical Quartely*, 40 (158), 1-12.

Morton, A. 1996: Folk psychology is not a predictive device. *Mind*, 105 (417), 119-137.

Morton, A. 2003: *The Importance of Being Understood: Folk Psychology as Ethics*. London: Routlegde.

Nelson, K. 2005: Language pathways into the community of minds. In Astington, J. and Baird, J.A., *Why Language Matters for Theory of Mind*. New York: Oxford University Press, 26-49.

Nelson, K., Plesa, D., Goldman, S., Henseler, S., Presler, N. and Walkenfeld, F.F. 2003: Entering a community of minds: an experiential aproach to 'theory of mind.' *Human Development*, 64, 24-46.

Nichols, S., Stich, S., Leslie, A. and Klein, D. 1996: Varieties of off-line simulation. In Carruthers, P. and Smith, P.K. (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 39-74.

Nichols, S. and Stich, S. 2003: *Mindreading: An Integrated Account of Pretence, Self-Awareness and Understanding of Other Minds*. Oxford: Oxford University Press.

Onishi, K.H. and Baillargeon, R. 2005: Do 15-month-old infants understand false beliefs? *Science*,308, 255-258.

Noë, A. 2004: *Action in perception*. Cambridge, MA: MIT Press.

Perner, J. 1996: Simulation as explicitation of prediction-implicit knowledge about the mind: arguments for a simulation-theory mix. In Carruthers, P. and Smith, P.K. (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 90-104.

Perner, J. and Ruffman, T. 2005: Infants insight into the mind: how deep? *Science*, 308 (5719), 214-216.

Povinelli, D.J. 2001: On the possibility of detecting intentions prior to understanding them. In Malle, B.F., Moses, L., and Baldwin, D.A. (eds.), *Intentions and Intentionality*. Cambridge MA: MIT Press, 225-248.

Povinelli, D.J. and Vonk, J. 2004: We don't need a microscope to explore the chimpanzee's mind. *Mind and Language*, 19 (1), 1-28.

Premack, D. and G. Woodruff. 1978: Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences,* 1, 515-26.

Ratcliffe, M. 2007: *Rethinking Commonsense Psychology: a critique of Folk Psychology, Theory of Mind and Simulation*. Basingstoke: Palgrave Macmillan.

Scheler, M. 1954. *The nature of sympathy*. Trans. P. Heath. London: Routledge & Kegan Paul

Slors, M.V.P. 2007: Intentional Systems Theory, Mental Causation and Empathic Resonance. *Erkenntnis* 67: 321-336.

Slors, M.V.P. 2010: Neural Resonance: Neither Simulation, Nor Percetion. *Phenomenology and the Cognitive Sciences* 9 (3): 437-458.

Southgate, V., Senju, A. and Csibra, G. 2007: Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587-592.

Spaulding, S. 2010: Embodied cognition and mindreading. *Mind and Language*, 25 (1), 119-140.

Surian, L., Caldi, S. and Sperber, D. 2007: Attribution of beliefs by 13-month-old infants.

Tomasello, M., Carpemter, M., Call, J., Behne, T. and Moll, H. 2005: Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675-735.

Wellman, H.M., Cross, D and Watson, J. 2001: Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655-684.

Wimmer, H. and Perner, J. 1983: Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 53, 45-57.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. 1995: An internal model for sensorimotor integration. *Science*, 269(5232): 1880–1882.

Zahavi, D. 2005: *Subjectivity and Selfhood*. Cambridge MA: MIT Press.

Zahavi, D.D. 2007: Expression and empathy. In Hutto, D.D. and Ratcliffe, M. 2007: *FolkPsychology Reassessed*. Dordrecht: Springer, 25-40.

Zawidski, T. 2008: The function of folk-psychology: mind-reading or mind-shaping? *Philosophical Explorations*, 11 (3), 193-217.